Summer 2014

# Novel Bacterial Lineages in the Uncultured Candidate Division SR1

Sumreet Kaur Ghotra
*San Jose State University*

# NOVEL BACTERIAL LINEAGES IN THE UNCULTURED CANDIDATE DIVISION SR1

A Thesis

Presented to

The Faculty of Department of Biological Sciences

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Masters of Science

by

Sumreet Kaur Ghotra

August 2014

# NOVEL BACTERIAL LINEAGES IN THE UNCULTURED CANDIDATE DIVISION SR1

by

Sumreet Kaur Ghotra

APPROVED FOR THE DEPARTMENT OF BIOLOGICAL SCIENCES

SAN JOSÉ STATE UNIVERSTY

August 2014

| | |
|---|---|
| Dr. Cleber Ouverney | Department of Biological Sciences |
| Dr. Miri VanHoven | Department of Biological Sciences |
| Dr. Rachael French | Department of Biological Sciences |

ABSTRACT

NOVEL BACTERIAL LINEAGES IN THE UNCULTURED

CANDIDATE DIVISION SR1

by

Sumreet Kaur Ghotra

Complex and naturally occurring microbial communities are known to be predominantly composed of uncultured bacteria and archaea. A few representatives of uncultured prokaryotes (e.g., from the candidate divisions TM7 and SR1) have been found in environmental samples and in association with animals, including humans. It is not yet clear, however, how uncultured bacteria found at environmental sites relate phylogenetically to their counterparts associated with animal hosts. In this study, we investigated the diversity of the SR1 candidate division to better understand the divide between animal-associated and environmental lineages. We generated 16S rDNA gene clone libraries from samples obtained from the following two habitats: outflow from water in a sulfur cave and spring water sediment in open air. Out of the 421 SR1 clones sequenced, we identified 18 operational taxonomic units (OTUs) that were used in combination with SR1 reference sequences to build an SR1 phylogenetic tree. Phylogenetic analyses indicated that SR1 members associated with animal hosts clustered separately from those collected at the environmental sites. Further in-depth investigation may help to better characterize these bacteria and establish their nutritional requirements

and physiological role in their respective communities. In addition, the microbes associated with host animals that are easy to maintain in the laboratory (e.g., termites, cow rumen) or found in habitats easy to access may serve as model organisms for studying the mechanisms by which these bacteria interact with their respective hosts, especially if the sequence homology observed for 16S rDNA to other genes, such as those involved in pathogenicity or mutualism.

# ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Dr Cleber Ouverney for all his support and guidance during my research. I feel so fortunate that he believed in my potential and gave me an opportunity to become part of his research group. Special thanks to my committee members Dr Miri VanHoven and Dr Rachael French for their valuable suggestions and reviews that motivated me to perform my best in my research. This research would not have been possible without the support of many of the peer graduate students with whom I had the privilege to work in last 2 years. I am really thankful to Adam Caldwell for helping me conduct many experiments and in the better understanding of phylogenetic analysis. I would like to thank other peer graduate students Amelia Lindsey and Edith Nyako for their help and suggestions during the course of my research. I owe a special thanks to the National Science Foundation under NSF award number SC3GM082291

My work as a research student would have been incomplete without the support of my loving husband Tarun Banka who has been a doctoral student himself and understands the challenges of being in research. His support and guidance has been the key motivating factor behind my work. And as they say all the good things come in the end, I would like to thank my adorable son Vihaan who has filled my life with immense love and happiness. His bright smile gives me strength to achieve my goals in life and believe in myself.

# Table of Contents

# List of Figures

# List of Tables

# The Uncultivable Bacteria

Historically, bacterial identification has been based on the visible colonies formed by bacteria under lab conditions. The morphological, physiological, and biochemical traits exhibited by these microorganisms were widely used for classification and quantification. It was not yet known that these cultivable bacteria represent only a minor fraction of our widespread bacterial diversity [1,2].

Most naturally occurring bacteria do not form visible colonies on a nutrient medium under lab conditions. Previous researches indicate that more than 99% of the bacteria in our environment are uncultivable under lab conditions [2,3]. The observation famously known as the "Great Plate Count Anomaly" also suggests that the direct microscopic count of an environmental sample far exceeds the number of microorganisms capable of forming visible colonies under standard lab conditions [2]. Many hypotheses have been postulated explaining the uncultivable nature of these bacteria but very little is actually known about their role in the environment. Some researchers suggest these bacteria remain uncultured to this date due to their complex nutritional requirements, which are intricately associated within their natural environments and have not yet been replicated under lab conditions [4,6]. Microbial biofilm is one such complex biological community consisting of millions of different bacterial species found in close collaborations with each other [5]. It is now known that the metabolic byproducts of one microbial community act as raw material for the other microbial community [4,7]. Some of these compounds are so complex that their chemistry is still not known. A study conducted by Jon Clardy [4] demonstrated that a

16S rRNA clone distantly related to *Verrucomicrobia,* grows only in the presence of a siderophore from a particular neighbor microorganism.

There have been many cases where bacteria previously considered uncultivable were able to form viable colonies by adding a specific growth factor to the nutrient medium [7,8]. Giovanonni [8] were able to isolate the members of the SAR11 clade, which were previously considered uncultivable, using high throughput procedures for isolating cell cultures by diluting the natural microbial communities into very low nutrient medium. Acidobacteria is another group of bacteria, which was previously considered uncultivable and was finally cultivated using specific secondary metabolites in the culture medium [7].

The uncultivable microbial consortia, though unidentified and unknown, have been speculated to be involved in crucial functions in our ecosystems that directly impact our environment [10] and health [9]. They have been reported to be involved in various biogeochemical transformations in the environment [10], form a major bulk of human microbiota [9,11,12], and sustain complex microbial mat ecosystems [5]. An uncultivable bacterium, *Rochalimaea henselae,* has been identified as the causative agent of bacillary angiomatosis [13]. Another uncultured bacterium, TM7, has been found to be associated with human periodontitis [12,14]. The ecological role of uncultured bacteria is still an area of extensive exploration and analysis.

Carl Woese pioneered the use of small subunit RNA molecules for identification and quantification of uncultured bacteria and archaea [2,15]. Such studies performed on isolates obtained directly from natural environments were used to establish molecular phylogenies of the uncultured representatives and by the year 1993 numerous

environmental sequences representing these uncultivable bacteria started appearing in the GenBank database [3,16]. With recent advances in molecular approaches towards identification and classification of microorganisms, many species of uncultivable bacteria, previously unknown, have been identified and are now represented as rRNA gene sequences across the gene databases [2,3]. Because of its length of ~1500 bp (bp=base pair), ideal for sequencing and its universal distribution, the 16S ribosomal RNA subunit is most widely used for identification and quantification of uncultivable bacteria [15].

Using these sequence-based methods, a wide variety of uncultured bacteria are being actively investigated for their phylogeny, gene functions, and ecological behavior [2,15]. For instance, Podar et al. [18] used flow cytometry to separate cells belonging to members of the TM7 phylum followed by amplification and sequencing to expand their phylogenetic characterization. TM7 has no cultivable representative isolated under lab conditions and by targeting such members directly from our environment by reconstructing the complete genomes of these uncultivable bacteria, valuable information can be inferred regarding their occurrence and their possible role in the communities they thrive in [12,14].

# A Literature Review on SR1

Among a plethora of uncultivable microorganisms identified solely by their small ribosomal subunit is candidate division SR1. SR1 is a recently discovered group in the Domain Bacteria with ongoing research on its characterization and quantification in the environment using culture-independent, gene based studies. Members of SR1 were first encountered along with various other bacterial species in the hydrocarbon-contaminated aquifer in the Wurtsmith Air Force Base, in Iosco County, Michigan [19,20]. Due to lack of information, they were initially classified into another candidate division, known as OP11 [19,20], which was represented by uncultured members obtained from the Obsidian Pool, a Yellowstone National Park hot spring. Soon a number of SR1 sequences were deposited in public genomic databases by independent researchers carrying out microbial studies in diverse environments such as hydrothermal vents [21], deep-sea sediments [22], mesophilic sulfur and sulfide rich springs [10], termite gut [23,24], and diseased coral tissue [19].

In 2004, Harris et al. [20] further resolved OP11 into new candidate divisions and the OP11-4 sub-division was renamed as SR1 on the basis of the sequences obtained from the Sulfur River, Kentucky. Further in 2009, Davis et al. [19] identified multiple novel members of SR1 and expanded the characterization of members of SR1 to a large extent from diverse habitats using 16S rRNA Analysis. Some of these habitats included: sediments from Zodletone Spring, a sulfur-rich spring in southwestern Oklahoma, microbial mats obtained from Sperm Pool, a high-temperature, low-pH pool (55°C, pH 2.5) in Yellowstone National Park, fresh bovine ruminal contents and anaerobic freshwater pond sediments (Duck Pond) in Norman, Oklahoma [19]. Using fluorescent *in*

*situ* hybridization (FISH), Davis et al. [19] identified two morphotypes of SR1, bacilli and filamentous, from the Duck Pond (OK), Zodletone Spring (OK) and Sperm Pool (Yellowstone National Park) using SR1 specific fluorescent-labeled probes. A quantitative PCR (qPCR) analysis was also performed by Davis et al. [19] to quantify the members of SR1 from various diverse environments and reported the lowest 16S rRNA gene copy numbers in Duck Pond (0.009%) followed by bovine rumen (0.08%) and much higher numbers in Zodletone Spring sediments (11.6%) and Sperm Pool (48.7%) microbial mat samples. However, it is not known yet which environmental factors influence SR1 abundance in one environment over the other, except for the conspicuous presence of higher sulfide and sulfur levels in the latter two habitats. [19]. Members of SR1 have been found in significant numbers in sulfur rich environments, which emphasizes the idea that they could be involved in biogeochemical cycling of nutrients like sulfur within their ecosystems [19,20].

Presently, the candidate division SR1 is classified into 2 subdivisions: BH1, based on the sequences obtained from Japan deep-sea sediments, and BD2-14, based on sequences from the Black Pool, Yellowstone National Park. All SR1 sequences found in environmental and human samples fall within those two subdivisions. BD2-14 is more diverse and widely distributed as compared to BH1 [19]. BD2-14 is subdivided into 9 subgroups (named with Roman numerals I-IX) with members belonging to geothermal environments as well as low-temperature terrestrial and marine environments [19]. Members of the BH-1 lineage, on the other hand, have been encountered only in extreme environmental habitats such as the hot springs at Yellowstone National Park [19]. Davis et al. [19] also reported that subgroup III of BD2-14 was composed exclusively of the

SR1 clones retrieved from human (oral cavity [9], esophagus [25]), mammalian (cow rumen [26], rhinoceros feces [27]), or insect (termite) origins [23,24]. Most of the members of the other subgroups were identified to be environment-based sequences with the exception of few human and animal based sequences.

SR1 is a relatively rare group of bacteria where no members have been reported from the extensive studies on multiple soil samples. Also, there are no studies reported on the mode of nutrition of this culture-elusive group of bacteria [19]. In this study we attempted to further expand our understanding of the diversity of SR1 in the environment and better characterize its members based on the 16SrRNA, identifying many novel sequences and proposing new subgroups of SR1.

# Materials and Methods

## Sample sources

Alum Rock Regional Park (N 37° 23.831, W 121° 47.860) is located in San José CA and was used as the primary sampling site for this study. This historic park is known for its sulfurous mineral spring water and mineral water caves. We selected this site because SR1 16S rDNA has previously been isolated from sulfur-rich environments. Two samples were collected from this sampling site. One sample was collected from an outflow with pungent sulfurous smell inside a dark cave and was named AR1A with water temperature 26°C and pH 6.5 (Figure 1). The other sample was collected from the sediment of an open-air mineral spring pool and was named AR2A with water temperature 23°C and pH 6.0 (Figure 1).

## Sample preparation

Both water samples were collected in sterile 500 ml Whirl-Pak bags (Nasco, Fort Atkinson, WI) under aseptic conditions and were immediately transported on ice to the lab, which is about 30 min away. Once at the lab, samples were immediately homogenized by thorough mixing and a 50 ml aliquot from each sample was transferred into sterile centrifuge tubes (Fisher Scientific, Waltham, Massachusetts). Samples were then centrifuged within a span of 1 h of sampling at 9,000 X g for 15 min at 4°C in a Sorvall RC 5C centrifuge. The supernatant was discarded and 0.25 g of the pellet was weighed in a Mettler Toledo AB54-S balance (Mettler Toledo, Columbus, Ohio) and used for genomic DNA extraction.

**Extraction of genomic DNA**

Genomic DNA was extracted from 0.25g pellet of suspended particles in the Alum Rock spring and cave water samples following a modified bead-beater half-lysis protocol as detailed in Roh et al. [28]. Samples were incubated for 30 min at 65°C with 10 $\mu$l of 10% Triton-X and 2.5 $\mu$l of 10mg/ml Proteinase K which act as cellular denaturants. To this mixture, 200 ml of lysis buffer (100 mM Tris-HCl pH 7.4, 20 mM EDTA pH 7.2, 5M guanidine thiocyanate) was added and vortexed. One half of the sample was then transferred to tubes containing a mixture of 0.1 mm and 0.5 mm diameter of sterile silica beads. The tubes containing beads were vortexed at 5,000 rpm for 30 s in order to shear the genomic DNA [28]. Breaking up the DNA reduces supercoiling which can inhibit PCR amplification. The other half of the sample was left with intact genomic DNA in order to maximize the diversity and prevent loss of genes of interest during the shearing step. The sheared and intact DNA samples were recombined and 400 $\mu$l of 99% benzyl alcohol was added to separate DNA from proteins and lipids [28]. DNA was precipitated with 20 $\mu$l of 3M sodium acetate and 400 $\mu$l of 100% ethanol. It was repeatedly washed with 500 ml of 70% ethanol solution [28]. The DNA pellet was resuspended in a low salt-buffer solution [28]. Mechanical bead beating gave the optimal combination of high DNA yield, large DNA size and minimal humic acid contamination for PCR amplification [28,29]. The nucleic acid concentration and purity of each of the genomic DNA samples used in this study were measured using a Nanodrop ND-1000 spectrophotometer (ND-1000 version 3.1.2 software; Thermo Scientific, Wilmington, DE). The molecular size of the DNA was determined via agarose gel (0.8%)

electrophoresis and the extracted nucleic acid was stored at -20°C immediately following extraction.

**16S rDNA PCR amplification**

In this study we used the 16S rDNA marker gene to identify SR1 bacteria in our samples. A number of general bacterial primers and SR1 specific primers were used to identify SR1 and are listed in Table 1. The extracted genomic DNA samples were first screened for the presence of SR1 bacteria by PCR amplification with a broad range general bacteria primer Bac 8F and SR1 specific primers SR1-918R and SR1-1380R using Paq DNA polymerase (Table 1). From the screened samples, 1/100$^{th}$ dilution (in sterile MilliQ water) of each genomic DNA sample was used for PCR amplification of the 16S rDNA gene as it had the best PCR yield which was 13.9 ng/$\mu$l for AR1A and 13 ng/$\mu$l for AR2A. Each 12.5 $\mu$l PCR reaction consisted of, 1x PCR Buffer B (Fisher Scientific, Waltham, Massachusetts), 1.5 mM MgCl2, 0.2 mM of each deoxynucleoside triphosphate (dNTP), 0.2 pmol of each forward and reverse primer, 1.25x10-2 units of Stratagene Paq5000 DNA polymerase (Applied Biosystems, Foster City, California), and 3-30 ng of nucleic acid. PCR thermal cycling was carried out in an Applied Biosystems 2720 Thermal Cycler (Applied Biosystems, Foster City, California). PCR screenings were performed using Stratagene Paq5000™ DNA polymerase under the following conditions: initial denaturation at 95°C for 2 min, 35 cycles of denaturation at 95°C for 20 s, annealing temperature for 20 s, and extension for 30 s. A final extension time followed at 72°C for 5 min. The expected PCR product size was about 1300 bp with primer pair Bac 8F/ SR1-1380R and about 920 bp from the primer pair Bac 8F/SR1-918R. The PCR products were confirmed for their appropriate size on a 0.8% agarose gel

(International Biotechnologies Inc., New Haven, Connecticut) containing 0.04% Ethidium Bromide and run in a Horizon 58 gel electrophoresis system (Invitrogen, Carlsbad, California) using a FB300 electrophoresis unit (Fisher Scientific, Waltham, Massachusetts) at 80 V with 400 mA of current for 25 min. Gel images were visualized using a Bio-Rad Universal Hood II UV Illuminator (Bio-Rad, Hercules, California) and gel images were processed and annotated with the software Quantity One (version 4.6; Bio-Rad, Hercules, California).

The positively screened SR1 genomic DNA samples were further amplified using PCR in a 50$\mu$l reaction volume using the same primer combinations and cycling conditions described above. Promega GoTaq® Flexi DNA polymerase (Promega, Sunnyvale, California) was used for these screenings due to its increased sensitivity and higher fidelity as compared to Paq DNA polymerase. The amplified PCR products were purified and recovered using an E-Gel® CloneWell 0.8% SYBR® Safe gel and iBase™ Combo (Invitrogen, Carlsbad, California) [30] per the manufacturer's instructions.

**Table 1**. The complete list of primers with their respective sequence, function, and PCR annealing temperatures used in this study.

| **Primer** | **Sequence (5' – 3')** | **Function** | **PCR (°C)[a]** | **References** |
|---|---|---|---|---|
| Gen.Bac-8F | AGAGTTTGATCCTGGCTCAG | Cloning | 60 | 40 |
| SR1-918R | TGTGTTCCCCCGCCTATC | Cloning | 60 | 17 |
| SR1-1380R | TGAGTGCAAGGAACAGGG | Cloning | 60 | 19 |
| M13 Forward | GTAAAACGACGGCCAG | Sequencing | 55 | 41 |
| M13 Reverse | CAGGAAACAGCTATGAC | Sequencing | 55 | 41 |

**16S rDNA cloning and sequencing**

The extracted and purified PCR products were ligated into pCR®2.1-TOPO vectors (Invitrogen, Carlsbad, California)[30] overnight at 14°C and transformed into One Shot TOP10 competent cells per the manufacturer's instructions [30]. The transformed cells were plated onto LB plates, each containing 50 $\mu$g/ml kanamycin and 40 mg/ml X-gal, for white-blue screening. This differential procedure is used to separate white colonies (have the DNA vector insert) from the blue colonies (lack the proper insert). For this, the plates were incubated at 37°C for 14-16 h. White colonies were transferred onto a new LB plates. Colony lysates were prepared to obtain DNA from the selected white colonies (clones) by transferring a small amount of the colony in 100 $\mu$l of sterile water and incubating for 10 min at 95°C in a 2720 thermo cycler (Applied Biosystems, Foster City, California) for 10 min at 95°C. This step enables the lysis of the cells and transfer of the insert into the solution.

The clones were PCR screened for the proper insert size using the same PCR conditions described above. PCR primers M13Forward and M13Reverse (Table 1) were used for library screening. Clones with the correct insert size were prepared for plasmid extraction using the QIAprep® Spin Miniprep Kit [31]. Clones were sent for Sanger sequencing at Sequetech Corporation (Mountain View, California) and were also stored in glycerol stocks at -80°C for future investigations.

**DNA sequence analysis**

The 16S rDNA sequences were imported into CodonCode Aligner (version 3.0.2; CodonCode Corporation, Dedham, Massachusetts) to trim the primer/vector sequence and poor-quality end regions. An open-source program, Mothur (version 1.18.1;

Department of Microbiology & Immunology, University of Michigan [http://www.mothur.org]) containing the Pintail and Chimera Slayer algorithms were used to check for chimeras in all sequences. No chimera sequence was found. All cleaned and chimera-checked sequences were compared for homology to other 16S rDNA sequences in the NCBI nucleotide database using Basic Local Alignment Search tool (BLAST) [32]. A Python script in BioPython (version 1.5.7; Cock et al. [http://biopython.org]) was used to conduct this search [33]. All BLAST hits with more than 98.5% similarity to other SR1 sequences were selected (Table 3). Sequences with no identity to any other SR1-related sequences in BLAST were removed from further analysis. Operational taxonomic unit (OTU) identification was done using the software USEARCH (version 4.1.93; R. Edgar [http://www.drive5.com/usearch])[34]. The clones more than 98.5% similar to each other were clustered into a single OTU. Clones less than 98.5% similar to any other SR1 OTU were identified as a unique sequence. The unique sequences were then sequenced from both ends and a contig for each gene was assembled using CodonCode Aligner assembler function (CodonCode Corporation, version 3.0.2, Dedham, Massachusetts).

**SR1 phylogenetic tree**

The cleaned sequences and their respective top BLAST hits were aligned using the online aligner SILVA (release SSU r106; Pruesse et al. [http://www.arb-silva.de]) [35]. The aligned sequences were imported into the tree-building software tool ARB (version 5.2; Ludwig et al. [http://www.arb-home.de/]) with the SILVA SSU Ref NR 106 database [36]. An SR1 group-specific filter was generated using an ARB filter tool. This filter contained 733 columns of unambiguous bases to manually clean up the clones.

Clones less than 733 base pairs were not included in the tree building and further analyses. A phylogenetic tree was constructed with the final 421 aligned SR1 sequences, the SR1 clones from the SILVA database, the BLAST top hits, and the SR1 subgroups proposed by Davis et al [19]. The tree was generated using the software PHYLIP (version 3.69, Department of Genome Sciences, University of Washington [http://evolution.genetics.washington.edu/phylip.html]), which uses a neighbor-joining algorithm to generate the tree based on a 1000 replicates for the bootstrap analysis to test the strength of the tree-branches (Figure 4). *E.coli* clone number X80725 was used to root the tree [37]. The consensus tree was then exported from ARB and imported into jsPhyloSVG ([http://www.jsphylosvg.com/]) to generate an interactive tree version (http://www.phylotouch.com/SR1) (Figure 4) [38].

**Rarefaction**

Rarefaction curves were generated for the two clone libraries with more than 30 SR1 clones for each clone library using BiodiversityR (version 1.5, World Agroforestry Centre [http://www.worldagroforestry.org/resources/databases/tree-diversity-analysis]). A Python script was used to group each of the SR1 sequences into their respective clone libraries and each clone library was further split into groups of 10 clones. The AR1A library had 151 clones and was split into groups of 10 clones each. The AR2A clone library had 270 clones and was split into 27 groups of 10 clones each (Figure 5). BiodiversityR was also used to calculate abundance (total number of clones per library) and richness (unique phylotypes, based on the OTUs) values for each clone library (Table 2).

**GenBank**

The 13 unique OTU seed sequences identified in this work have been deposited in the GenBank database with accession numbers JF917269 to JF917281, as shown in the Table 3 [16,39].

# Results

**Sample collection and DNA extractions**

Both of the sampling sites identified in this study, the mineral spring water sediments from an open pool and sediments from the sulfur cave outflow (Figure 1), proved to be successful sites for identification of SR1 bacteria. The modified bead beater protocol used in this study resulted in fairly high yields of DNA.


**Figure 1**. The two sampling sites used in this study. Site 1: The sulfur cave from which sample AR1A was collected from water sediment. Site 2: The open-air mineral spring from which sample AR2A was collected from sediment.

**PCR screenings**

Two SR1-specific primer-pair combinations were used to screen the extracted genomic DNA for the presence of SR1 using PCR: Bac 8F/SR1-918R and Bac 8F/SR1-1380R. The primer-pair combination Bac 8F/SR1-918R was used to screen the sulfur cave water sample AR1A, whereas primer-pair Bac 8F/ SR1-1380R was used to screen both the water samples from the sulfur cave (AR1A) as well as the mineral spring (AR2A). The expected PCR product for Bac 8F/ SR1-1380R was about 1300 bp, while that from the primer-pair Bac-8F/SR1-918R was about 920 bp. Also, the primer-pair combination Bac 8F/1380R generated higher number of clones as compared to the primer-pair Bac 8F/SR1-918R and was used for all the further analyses.

The gel image shown in Figure 2 indicates a bright positive band from the sample AR2A at 1/100th dilution using primer-pair combination Bac 8F/ SR1-1380R. Similar positive bands were obtained for the sample AR1A using the two primer-pair combinations (data not shown). This confirmed the presence of SR1 in the water samples collected from the sulfur cave and mineral spring.

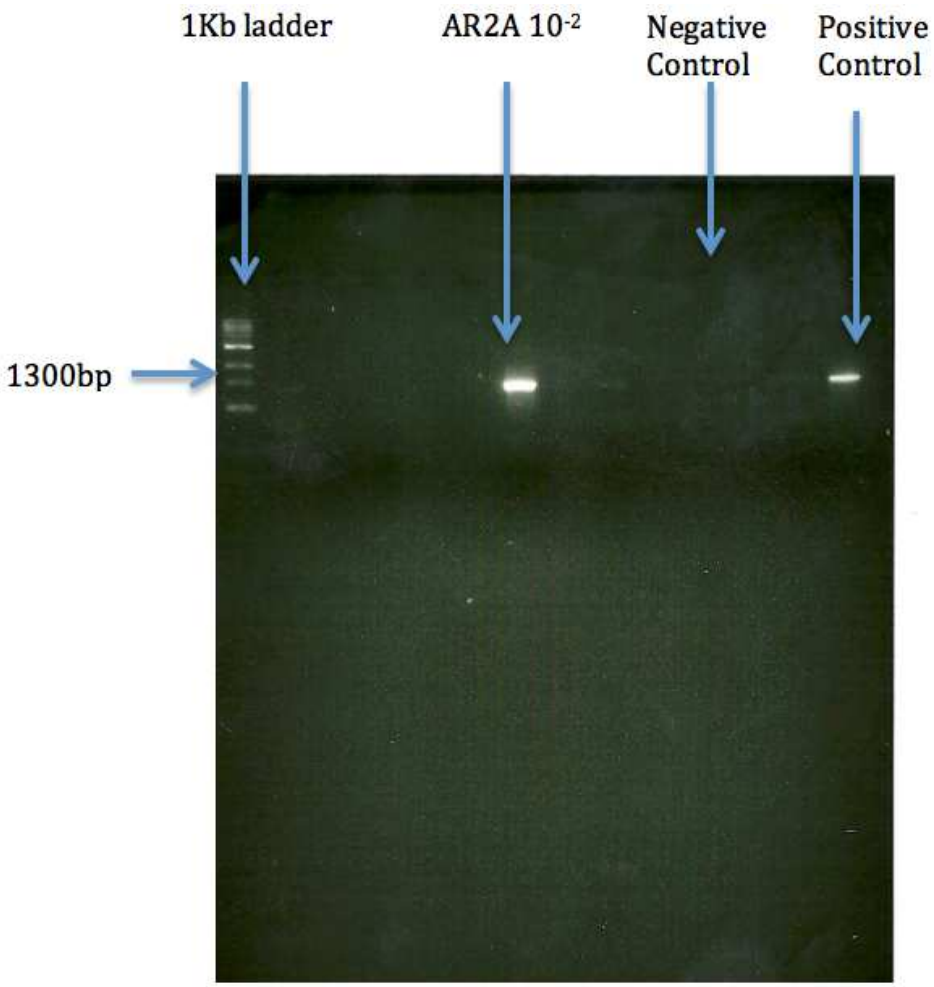


**Figure 2.** Agarose gel images depicting PCR results with primers Bac 8F and SR1-1380R used for screening genomic DNA extracted with modified bead-beater protocol. The 1Kb molecular weight (ladder) was used to identify the size of the sample. The sample AR2A, shown in this image, is visible at the 1300 bp marker. No negative control showed up during screening. Positive control is clearly visible in this image. AR1A is not shown in this image.

**Clone libraries**

Three 16S rDNA PCR clone libraries were generated in this study. Two libraries were generated from the cave water sample AR1A using the two primer-pair combinations Bac 8F/SR1-1380R and Bac 8F/SR1-918R. One library was generated from the spring water sample AR2A using the primer-pair combination Bac 8F/SR1-1380R.

From a total of 421 clones, 151 clones were generated from the cave water sample AR1A. Of these 151 clones, 14 clones were generated using primer-pair Bac 8F and SR1-918R and 137 clones were generated using Bac 8F and SR1-1380R. In addition, 270 clones were generated from the spring water sample AR2A using the primer pair Bac 8F and SR1-1380R. Further sequencing of the clones from AR2A was discontinued as no unique phylotypes from the site were being generated as per the rarefaction analysis (Figure 5).

The most abundant phylotype in our study, AR1A-1380R-117, had a total of 259 clones and was detected in all three clone libraries. Out of the three clone libraries, the sample sources, abundance (number of clones), and richness (number of unique clones) from the two libraries are listed in (Table 2). The spring water sample AR2A was the most abundant and richest between the two clone libraries generating 270 clones (64%) out of the 421 clones.

**Table 2.** Summary of the two samples, the primer pairs used for cloning SR1 16S rDNA genes and the two clone libraries generated. The clone abundance and richness values were calculated using BiodiversityR as detailed in the methods.

| **Sample name** (Sample Source) | **PCR primer pair** | **Clone library** | **Abundance (%)** | **Richness** |
|---|---|---|---|---|
| **AR1A** (Sulfur Cave) | Bac-8F & SR1-918R<br>Bac-8F & SR1-1380R | AR1A-918R<br>AR1A-1380R | 14 (2.0)<br>137(20.2) | 3<br>12 |
| **AR2A** (Mineral Spring) | Bac-8F & SR1-1380R | AR2A-1380R | 270 (39.7) | 17 |

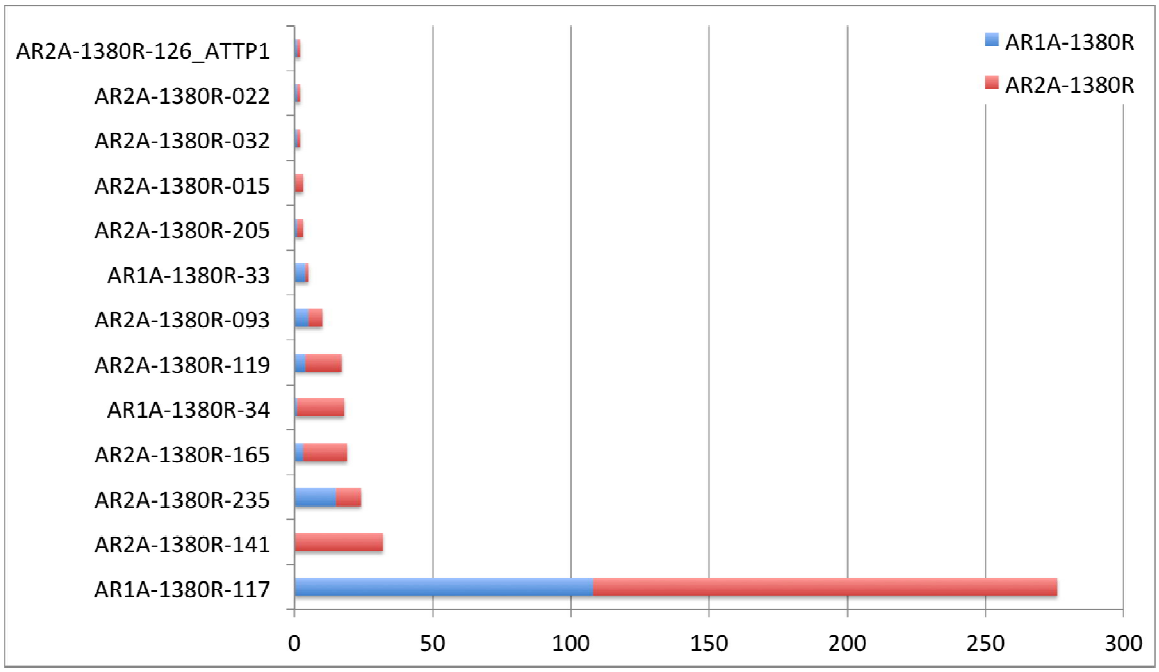


**Figure 3.** The two clone libraries and the most common phylotypes. AR1A-1380R-117 was found to be the most common phylotype with about 259 clones in its group.

**BLAST analysis**

The Basic Local Alignment Tool (BLAST) is a database tool provided by NCBI (National Center for Biotechnology Information) used to compare regions of local similarity between sequences. The program compares nucleotide sequences to the sequences in the database and calculates the statistical significance of the match between the sequences. This program was used to compare the sequences from this study to the available sequences in the GenBank database to infer the percentage similarity with the top hits found in the search results (Table 3). Out of the 18 unique OTU seed sequences from this study, 13 were deposited in GenBank, numbered from JF917269-JF917281 with their corresponding top hits from the BLAST search results. All the corresponding percentage similarity values are above 85%. SR1 clone number JF917281 is the only clone from this study that had a human-associated clone in the BLAST top hit results. JF917281 was 86% homologous to an SR1 in human stool (accession number JQ191085). All other clones from this study were closely related to other environmental SR1 clones as the respective top hits.

**Table 3**. The 13 unique OTUs have been deposited in GenBank, with accession numbers JF917269-JF917281 and their corresponding top hits from the BLAST search results with their respective similarities (in percentage).

| # | Accession Number | BLAST Top Hit | % Similarity |
|---|---|---|---|
| 1 | JF917269 | Terrestrial sulfidic spring cloneEPSO1BB | 96% |
| 2 | JF917270 | Terrestrial sulfidic spring clone JX521644 | 99% |
| 3 | JF917271 | Terrestrial sulfidic spring clone JX521155 | 99% |
| 4 | JF917272 | Terrestrial sulfidic spring clone JX521481 | 97% |
| 5 | JF917273 | Hypersaline Microbial mat clone JN439316 | 85% |
| 6 | JF917274 | Sulfidic cave water clone EU101267 | 87% |
| 7 | JF917275 | Sulfidic cave water clone EU101267 | 99% |
| 8 | JF917276 | Hypersaline Microbial mat clone JN491965 | 94% |
| 9 | JF917277 | Anaerobic Bioreactor clone GQ181297 | 98% |
| 10 | JF917278 | Hypersaline Microbial mat clone JF491965 | 94% |
| 11 | JF917279 | Sulfidic springs clone JF521644 | 86% |
| 12 | JF917280 | Sulfidic springs clone JF521644 | 95% |
| 13 | JF917281 | Human stool JQ191085 | 86% |

**Phylogenetic analysis**

A cutoff of 98.5% was used to identify unique OTUs. Using this standard, we identified 18 new unique OTUs from a total of 421 clones from both the sampling sites. The SR1 tree in this study was generated using 421 SR1 sequences consisting of our 18 unique OTUs, 95 SR1 reference sequences and *E. coli* X80725 from GenBank as the outgroup to root the tree. Almost all the branches were strongly supported with bootstrap values more than 50% by generating 1000 iterations of the tree (Figure 4).

The previously proposed SR1 tree has 9 subgroups. In this study a new subgroup, Subgroup X, was proposed on the basis of grouping of the sequences. Subgroup I is purely composed of environmental sequences from hydrothermal vent and microbial mat clones. Only one unique sequence from this study fell into this subgroup. Subgroup II is composed of both environmental as well as host-associated sequences. Host sequences are from rhinoceros, dolphin and rumen of the cow and environmental clones spring sediment and lake water. No clone from this study fell into subgroup II.

Subgroup III is a unique subgroup consisting purely of the host associated sequences mostly from human skin, esophagus and oral cavity, cow rumen, rhinoceros feces and termite gut. None of the environmental-based sequences from this study were found in this subgroup. In addition, subgroup III showed a notable internal split, with all 11 human-associated sequences clustering together except for clone EU382051, which originated in a cow rumen in Alberta, Canada. All other sequences in subgroup III are either from termite gut, cow rumen or rhinoceros feces, and they cluster separately from the human-associated sequences. Subgroup IV is composed of environmental sequences

and no host-associated sequences. Five of the unique phylotypes from this study fell into this subgroup. Subgroup V is also composed of environmental clones mostly from lake water, microbial mats, activated sludge and deep sea sediments. No clones from this study were found in this subgroup.

Subgroup VI is purely composed of environmental clones from spring sediment and lake water and six unique clones from this study fell into subgroup VI. Subgroup VII is mostly composed of environmental sequences barring few host-associated clones from dog duodenum and diseased coral tissues and two unique clones from this study fell into subgroup VII. Subgroup VIII is composed of sulfur river clones and one unique clone from this study fell into this subgroup. Subgroup IX is composed of spring water clones and no clones from this study. Subgroup X is the new proposed subgroup of the SR1 tree, exclusively composed of the clones from this study.

Clone number JF917281 from the proposed subgroup X was the only clone found to be 86% similar to the human stool clone JQ191085 from the BLAST results. This clone is the only one found to be associated with the human clone from this study. All the other clones were found to be associated with environmental clones.

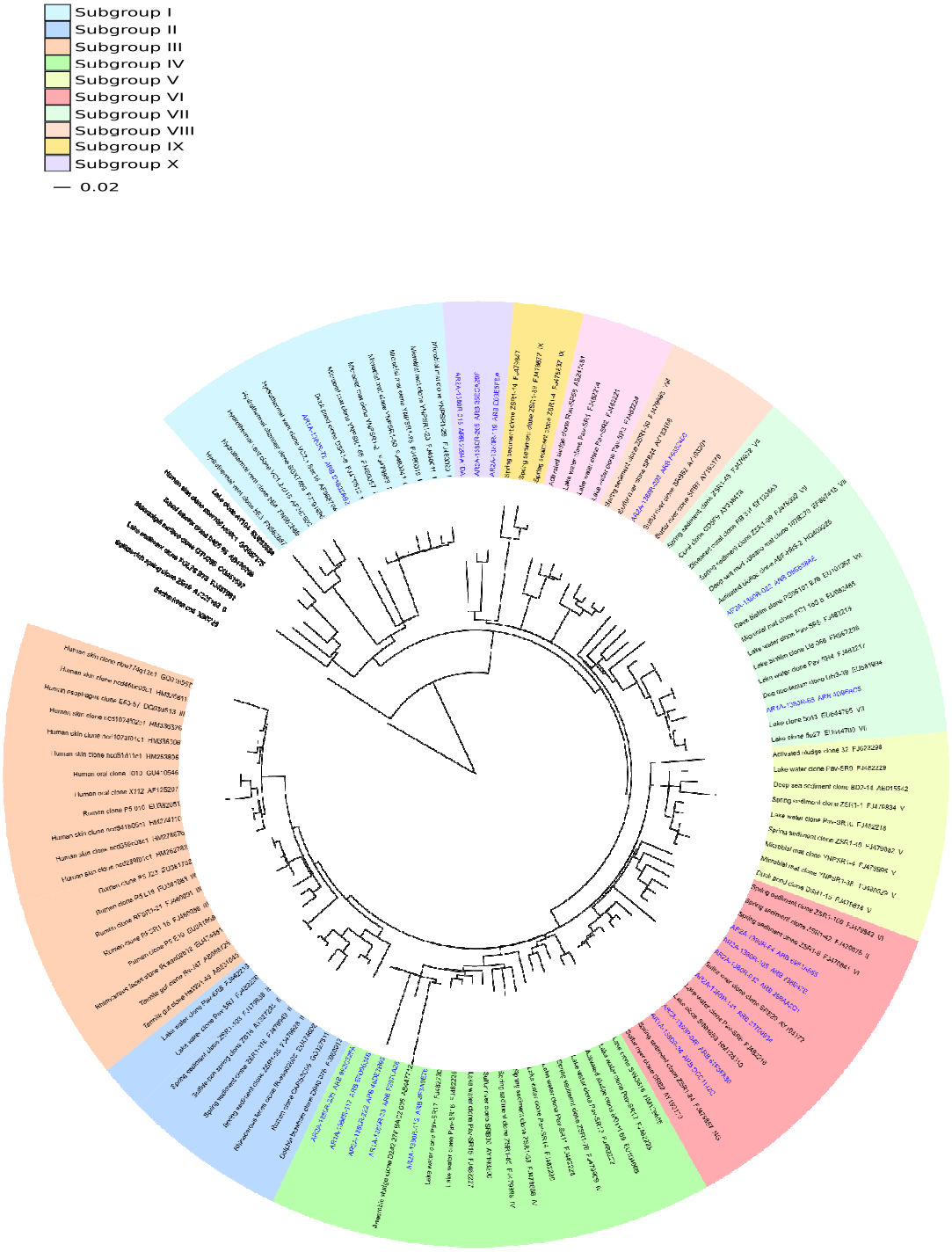


**Figure 4.** Phylogenetic tree of candidate division SR1 with 129 SR1 OTUs (>98.5 %

identity cut off), including the 18 unique OTUs from this study (blue font), plus outgroup *E. coli* GenBank accession number X80725. Bootstrap values >50% are shown. The outgroup branch length shown at 2/3 scale (original length 0.31389, shown length to 0.20926). An interactive version of this tree, built using jsPhyloSVG can be seen at www.phylotouch.com/sr1.

**Rarefaction analysis**

The rarefaction curves were generated for both the clone libraries in our study using the software BiodiversityR (Figure 5). AR2A is the most diverse clone library because its rarefaction curve lies above AR1A as it has higher number of unique phylotypes per clone sequenced. Both clone libraries have rarefaction curves with non-zero slopes at the highest abundance values indicating that further unique SR1 phylotypes could be found in these sites.

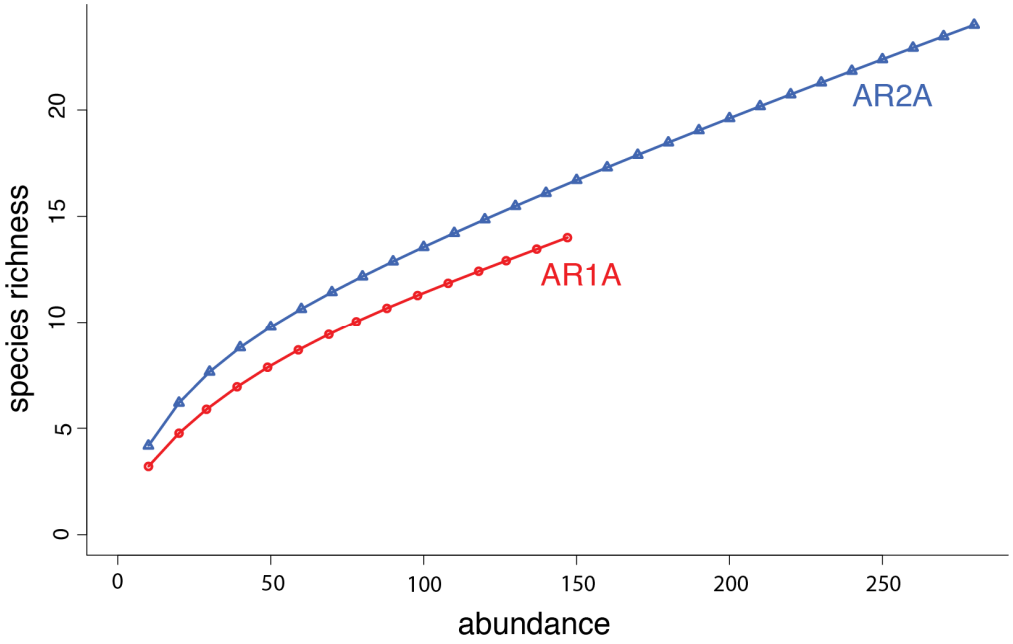

**Figure 5.** Rarefaction curves for both the clone libraries AR1A and AR2A using BiodiversityR. The red slope represents AR1A and the blue slope represents AR2A. The X-axis indicates abundance and the Y-axis represents the species richness.

# Discussion

Environmental sequences are being widely used as models to study host-associated sequences, as any investigation on human subjects is very challenging and restricted. Most published SR1 sequences are environmental in origin, except for those in Subgroup III from the latest SR1 tree published by Davis et al in (2009). They also established that Subgroup III is purely composed of host-associated clones that are phylogenetically very distant to the environmental-based clones. This unique subgroup is composed of the clones that are human-associated, mostly from the skin, oral cavity, and esophagus. The other host-associated clones are from the termite gut, rhinoceros feces, dolphin blowhole, and cow rumen. Among these hosts, termites are the only ones that can be easily maintained in the laboratory and can serve as model organisms for the 16S rDNA analysis. This has also been established in a study conducted by Dinis et al. (2011) where they found that the TM7 16S rDNA from the termite gut were closely associated to the human 16S rDNA sequences from the oral and vaginal cavities.

It is very clear from the SR1 phylogenetic analysis that the host associated sequences are very different from the clones found in the other environmental habitats such as the microbial mats, lake water, deep sea sediments, and spring water. Barring a few isolated host-associated clones such as the squat lobster clone, dog duodenum clone, and the human skin clone found scattered in other subgroups, the sequences in the rest of the subgroups are environment based. This suggests that the host associated clones have co-evolved with their respective hosts in such a way that they can survive only in the hosts and not in the other environments. But there are very limited published data to support this hypothesis. Extensive phylogenetic and genomic analysis is needed to further

investigate this relationship between the SR1 bacteria and their respective hosts to establish whether this relationship is mutualism, commensalism or has no effect on either of species.

Subgroup II is a transitional subgroup that is composed of both host associated and environmental clones. The host-associated clones are from dolphin, cow and rhinoceros while the environmental clones are from spring sediment and lake water. Subgroup X is the newly proposed subgroup from this study that has sequences exclusively from this study from mineral spring water samples. No sequences from this study were found in subgroups II, III, V and IX and further analysis is required to establish whether it is due to the environment-specific behavior of these sequences or due to incomplete cloning, PCR bias, PCR inhibition or some other factors. Also no clones from this study were found in subdivision BD2-14, which is composed of sequences mostly from the extreme habitats such as hydrothermal vents and deep-sea sediments. Also, there is a lack of published data on the environments from where these sequences were identified. For example, the sequences from the microbial mats could be actually from other organisms (bacteria/fungi) present in these complex environments. Similar discrepancies regarding the source of the sequences arise from the clones isolated from diseased coral tissues, activated wastewater and sediments from lake/spring.

This study was an extensive attempt to expand the characterization of the candidate division SR1 and help ascertain an established status for this as-yet uncultured group. The rarefaction analysis from this study also suggests that this group is still not fully characterized and it is crucial to completely resolve this poorly explored candidate division. Full genome sequencing of this group of bacteria may help to resolve the status

and the possible biogeochemical role of these bacteria in the environment. This may also aid in establishing their speculated role in sulfur cycling due to their exclusive presence in the sulfur rich environments. Their presence in diseased coral tissues and oral cavity may indicate their possible involvement in the diseases associated with such animal/human hosts. The sampling problems encountered due to extremely low numbers of these bacteria in their respective environments accounts for the limited information on these bacteria. This problem can also be addressed by full genome sequencing of the 16S rDNA, which is still the most widely accepted criteria to characterize and classify uncultured groups of bacteria.

Since the identification of the first SR1 sequences from the contaminated hydrocarbon aquifer, a plethora of SR1 sequences have been deposited in public databases and these investigations are still ongoing. This study adds further information to what is already known about this elusive group of bacteria and is another crucial step towards establishing its significance in our environment.

# Conclusion

Because of its low abundance, SR1 is an elusive group of microorganisms when compared to many other uncultured microbes. Regardless, it is crucial to further broaden the SR1 diversity and investigate its respective role in the environment. This study was successful in identifying many novel SR1 lineages and helped better understand the classification of members in the SR1 candidate division. Also some new habitats (Alum Rock Regional Park mineral spring and sulfur cave) were recognized as sources for isolation and identification of this bacterium. This is an important addition to the pre-existing investigation as this bacterium is rarely isolated from the environment and if so, in very low numbers. However, a deeper insight into this uncultured division of bacteria may shed light on its possible role in the environment. Full genome analysis of this elusive group of bacteria will be a critical milestone in completely classifying this group as well as identifying its gene-function that can shed light on its possible role in the environment and establish its ecological status.

# References


1. **Rappe, M. S., and S. J. Giovannoni.** 2003. The uncultured microbial majority. Annual review of microbiology **57:**369-394.

2. **Amann R., Ludwig, W. and Schleifer, K.H.** 1995. Phylogenetic identification and *in-situ* detection of individual microbial cells without cultivation. Microbiol. Rev., **59:** 143–169.

3. **Hugenholtz, P., B. M. Goebel, and N. R. Pace.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol **180:**4765-4774.

4. **D'Onofrio, A., Crawford, J. M., Stewart, E. J., Witt, K., Gavrish, E., Epstein, S., Clardy, J., and Lewis, K.** Siderophores from Neighboring Organisms Promote The Growth of Uncultured Bacteria.

5. **Lau, M. C., J. C. Aitchison, and S. B. Pointing.** 2009. Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. Extremophiles **13:**139-149.

6. **Dethlefsen, L., M. McFall-Ngai, and D. A. Relman.** 2007. An ecological and evolutionary perspective on human-microbe mutualism and disease. Nature **449:**811-818.

7. **Lewis K, Epstein S, D'Onofrio A, Ling LL.** 2010. Uncultured microorganisms as a source of secondary metabolites. The Journal of antibiotics **63:**468-476.

8. **Rappé MS, Connon SA, Vergin KL, Giovannoni SJ.** 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature **418:**630-633.

9. **Paster, B. J., S. K. Boches, J. L. Galvin, R. E. Ericson, C. N. Lau, V. A. Levanos, A. Sahasrabudhe, and F. E. Dewhirst.** 2001. Bacterial diversity in human subgingival plaque. Journal of Bacteriology **183:**3770-3783.

10. **Elshahed MS, Senko JM, Najar FZ, Kenton SM, Roe BA, Dewers TA, Spear JR, Krumholz LR.** 2003. Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring. Applied and Environmental Microbiology **69:**5609-5621.

11. **Bik, E. M., P. B. Eckburg, S. R. Gill, K. E. Nelson, E. A. Purdom, F. Francois, G. Perez-Perez, M. J. Blaser, and D. A. Relman.** 2006. Molecular analysis of the bacterial microbiota in the human stomach. Proc Natl Acad Sci U S A **103:**732-737.

12. **Ouverney CC, Armitage GC, Relman DA.** 2003. Single-Cell Enumeration of an Uncultivated TM7 Subgroup in the Human Subgingival Crevice. Applied and Environmental Microbiology **69:**6294-6298.

13. **Relman, D. A., Loutit, J.S., Schmidt, T. M., Falkow, S., and Tompkins, L.S.** 1990. The agent of bacillary angiomatosis: an approach to the identification of uncultured pathogens. N. Engl. J. Med. **323:**1573–1580.

14. **Dinis, J. M., D. E. Barton, J. Ghadiri, D. Surendar, K. Reddy, F. Velasquez, C. L. Chaffee, M. C. Lee, H. Gavrilova, H. Ozuna, S. A. Smits, and C. C. Ouverney.** 2011. In search of an uncultured human-associated TM7 bacterium in the environment. PLoS One **6:**e21280.

15. **Weng L, Rubin EM, Bristow J.** 2006. Application of sequence-based methods in human microbial ecology. Genome research **16:**316-322.

16. **Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler.** 2005. GenBank. Nucleic Acids Res **33:**D34-38.

17. **Hugenholtz, P.** 2011. Primer generated to target the SR1 candidate phylum and used with permission.

18. **Podar, M., C. B. Abulencia, M. Walcher, D. Hutchison, K. Zengler, J. A. Garcia, T. Holland, D. Cotton, L. Hauser, and M. Keller.** 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. Applied and environmental microbiology **73:**3205-3214.

19. **Davis, J. P., N. H. Youssef, and M. S. Elshahed.** 2009. Assessment of the diversity, abundance, and ecological distribution of members of candidate division SR1 reveals a high level of phylogenetic diversity but limited morphotypic diversity. Appl Environ Microbiol **75:**4139-4148.

20. **Harris, J. K., S. T. Kelley, and N. R. Pace.** 2004. New perspective on uncultured bacterial phylogenetic division OP11. Appl Environ Microbiol **70:**845-849.

21. **Reysenbach, A. L., K. Longnecker, and J. Kirshtein.** 2000. Novel bacterial and archaeal lineages from an in situ growth chamber deployed at a Mid-Atlantic Ridge hydrothermal vent. Appl Environ Microbiol **66:**3798-3806.

22. **Li, L. N., C. Kato, and K. Horikoshi.** 1999. Bacterial diversity in deep-sea sediments from different depths. Biodivers Conserv **8:**659-677.

23. **Hongoh, Y., M. Ohkuma, and T. Kudo.** 2003. Molecular analysis of bacterial microbiota in the gut of the termite Reticulitermes speratus (Isoptera; Rhinotermitidae). FEMS Microbiol Ecol **44:**231-242.

24. **Nakajima, H., Y. Hongoh, S. Noda, Y. Yoshida, R. Usami, T. Kudo, and M. Ohkuma.** 2006. Phylogenetic and morphological diversity of Bacteroidales members associated with the gut wall of termites. Bioscience, biotechnology, and biochemistry **70:**211-218.

25. **Pei, Z., E. J. Bini, L. Yang, M. Zhou, F. Francois, and M. J. Blaser.** 2004. Bacterial biota in the human distal esophagus. Proc Natl Acad Sci U S A **101:**4250-4255.

26. **Brulc, J. M., D. A. Antonopoulos, M. E. Miller, M. K. Wilson, A. C. Yannarell, E. A. Dinsdale, R. E. Edwards, E. D. Frank, J. B. Emerson, P. Wacklin, P. M. Coutinho, B. Henrissat, K. E. Nelson, and B. A. White.** 2009. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. Proc Natl Acad Sci U S A **106:**1948-1953.

27. **Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, R. Knight, and J. I. Gordon.** 2008. Evolution of mammals and their gut microbes. Science **320:**1647-1651.

28. **Roh C, Villatte F, Kim B-G, Schmid RD.** 2006. Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. Applied biochemistry and biotechnology **134:**97-112.

29. **Yeates C, Gillings M, Davison A, Altavilla N, Veal D.** 1997. PCR amplification of crude microbial DNA extracted from soil. Letters in Applied Microbiology **25:**303-307.

30. **Invitrogen.** 2006. TA Cloning Kit User Manual. Invitrogen, Carlsbad, CA.

31. **Qiagen.** 2006. QIAPrep Miniprep Handbook. Qiagen, Valencia, CA.

32. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389-3402.

33. **Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B.** 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics **25:**1422-1423.

34. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:**2460-2461.

35. **Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glockner.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res **35:**7188-7196.

36. **Ludwig W, Strunk O, Westram R, Richter L, Meier H, Buchner A, Lai T, Steppi S, Jobb G, Förster W.** 2004. ARB: a software environment for sequence data. Nucleic acids research **32:**1363-1371.

37. **Cilia V, Lafay B, Christen R.** 1996. Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. Molecular biology and evolution **13:**451-461.

38. **Smits SA, Ouverney CC.** 2010. jsPhyloSVG: a javascript library for visualizing interactive and vector-based phylogenetic trees on the web. PLoS ONE **5:**e12267.

39. **Caldwell, A.C., Ghotra, S., Ohene-Nyako, T.E., Lindsey, A.R.I., Nystrom, S.N., and Ouverney, C.C.** 2011. Phylogenetic analysis of candidate division SR1 reveals new subgroups. GenBank Sequence: http://www.ncbi.nlm.nih.gov/nuccore/JF917269,JF917270,JF917271,JF917272,JF917273,JF917274,JF917275,JF917276,JF917277,JF917278,JF917279,JF917280,JF917281

40. **Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, and M. L. Sogin.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci **82:**6955-6959.

41. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol **101:**20-78.

# Appendix

## Full Sequences of 18 OTUs

>AR1A-1380R-117

GAGTTTGATCCTGGCTCAGGATGAACGCTAGCGGAACGCCTAACACATGCAAGTCGAACGAGTTTTATTTTGTAGCAAT
ACAGAATAAAATTAGTGGCAAACGAGCGAGTAACGCGTACTTAATCTGCCCAAAAGTTGAGAACAACTATAGAGATGT
AGCTAATTCTCAATATGCTCACGATGGTGAAAATCCGATGTGAGGAAAGATTTATCGCTTTTGGAGGAGGGTGCGTCCT
ATCAGGTTGTTGGTGAGGTAATAGCTCACCAAGCCTATGACGGGTAGCCGGTCTGAGATGATGTCCGGCCGCGATGGA
ACTGAGATACGGTCCATACTCCTACGGGAGGCAGCAGTGGGGAATCTTGCACAATGGGCGAAAGCCTGATGCAGCGAT
TCCGCGTGGAGGATGACGCATTACGGTGTGTAAACTCCTTTTTTGGTGGAAGACGAATGACGGTACACCAAGAATAAG
GGACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTCCCAAGCGTTATCCGGATTTACTGGGCGTAAAGTGT
CCGTAGCCTGGTATATAAGTCTGTTTTCAAATGCCACGACTCAATCGTGGAAAGGGAATAGATACTGTATACCTAGAGA
ATATTGGAGGTTAGTGGAATTTCCGGTGGAGCGGTGAAATGCGTTGATATCGGAAAGAACGCCGAAAGCGAAAGCAGC
TAACTATGATATTTCTGACGGTGATGGACGACAGCTTGAGTAGCAAACGGGATTAGATACCCCGGTAGTTCAAGCCGTA
AATTATCCTTGCTAGGTGTCTCCATAACTATTGGTATAGCATTTGTTATATCAAAAGTTATGGAGGTGCCGTAAGTTAAC
ACGTTAAGCAAGGCGCCTGTGTAGTATATTCGCAAGAATGAAACACAAAGGGATAGGCGGGGGAACACACAAGCAGT
GGATGATCTAGATTAATTGGATACTAAGCCAAGAATCTTACCCAGACTTGACATGGATTGTGTCTACGATGAAAGTCGT
ATAGCTATAGCAATATAGCGCTTTTCACAGATGGTGCATGGTCGTCGTCAGCTCGTGCCGCAAGGTGTCTAGTTAAGTC
TAGTAACGAGCGCAACCCTCATCGTTAGTTAGTTATGTCTAACGAGACTGCCCGCGTAAGCGGGAGGAAGGAGAGGAT
GACGTCAGATCCTCATGCCTCTTACGTCCGGGGCTTCATAGATCATACAATGGGTAGTTACAATGAGAAGCAATACCGC
AAGGTGGAGCAAATCTTCAAAGCTACCCTCAGTTCGGATTGTAGTCTGGAATTCGACTACATGAAGTTGGAATTGCTAG
TAATGGTGAATCAGCCATGTCACCGTGAATATGTCCCTGTTCCTTGCACTCAAAGCCGAATTCCAGCACACTGGCGGCC
GTTACTAGTGGATCCGAGCTCGGTACCAAGCTTGATGCATAGCTTGAGTATTCTATAGGTCACC

>AR2A-1380R-235

GGAAGAGGGACGTATTCACGGTGACGTGGATGATTCACCATTACTAGCAATTCCTACTTCTTGTAGTCGAATTCCAGAC
TACAATCCGAACTGAAAGTAGCTCTGAAGATTTGCTCCACCTTGCGGTATTGCTTCTCATTGTAACTACCCCTTGTATGA
TCTATGAAGCCCCAGACGTAAGAGGCATGAGGATCTGACGTCATCCTCTCCTTCCTCCCGCTTACGCGGGCAGTCTCGT
TAGACATAACTAACTAACGATGAGGGTTGCGCTCGTTACTACACTTAACTAGACACCTTGCGGCACGAGCTGACGACG
ACCATGCACCATCTGTGAAAAGCGCTATATTGCTATAGCTATACGACTTTCATCGTAGACACAATCCATGTCAAGTCTG
GGTAAGATTCTTGGCTTANTATCCCATTAATCTAGATCATCCACTGCTTGTGTGTTCCCCCGCCTATCCCTTTGTGTTTCA
TTCTTGCGAATATACTACACAGGCGCCTTGCTTAACGTGGTAACTTACGGCACCTCCATAACTTTTGATATACCAATGCT
ATACCAATAGTTATGGAGACACCTAGCAAGGATAATTTACGGCTGAACTACCGGGGTATCTAATCCCGTTTGCTACTCA
AGCTGCGTCCATCACCGTCAGAAATATCATAGTTAGCTGCTTTCGCTTTCGGCGTTCTTTCCGATATCAACGCATTTCAC

CGCTCCACCGGAAATTCCACTAACCTCCAATATTCTCTAGGTATACAGTATCTATTCCCTTTCCACGATTGAGTCGTGGC
ATTTGAAAACAGACTTATATACCAGGCTACGGACACTTTACGCCCAGTAAATCCGGATAACGCTTGGGACCTACGTATT
ACCGCGGCTGCTGGCACGTAGTTAGCCGTCCCTTATTCTTGGTGTACCGTCATTCGTCTTCCACCAAAAAAGGAGTTTAC
ACACCGTAATGCGTCATCCTCCACGCGGAATCGCTGCATCAGGCTTTCGCCCATTGTGCAAGATTCCCCACTGCTGCCT
CCCGTAGGAGTATGGACCGTATCTCAGTTCCATCGCGGCCGGACATCATCTCAGACCGGCTACCCGTCATAGGCTTGGT
GAGCTATTACCTCACCAACAACCTGATAGGACGCACCCTCCTCCAAAAGCGATAAATCTTTCCTCACATCGGATTTTCA
CCATCGTGAGCATATTGAGAATTAGCTACATCTCTATAGTTGTTCTCAACTTTTGGGCAGATTAAGTACGCGTTACTCGC
TCGTTTGCCACTAATTTTATTCTGTATTGCTACAAAATAAAACTCGTTCGACTTGCATGTGTTAGGCGTTCCGCTAGCGT
TCATCCTGAGCCAGGATCAAACTCTAAGCCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCGGTACC
AAGCTTGATGCATAGCTTGAGTATTCTATAGNTCACCNANNNNNN

>AR2A-1380R-093

TGAGTGCAAGGAACAGGGACATATTCACGGCGACATGGCTGATTCGCCATTACTAGCAATTCCAACTTCATGTAATCAG
GTTCCAGATTACAATCCGAACTAGGGGTAGGTTTAAAGATTTGCTCCATCTCACGATATCGCTTCTCACTGTTGCTACCC
ATTGTATTGTCTATGAAGCCCTAGACATAAGGGGCATGAGGATCTGACGTCATCCTCCCCTTCCTCCCGGTTGCCCGGA
CAGTCTCTTTAGACATACTAACTAAAGATGAGGGTTGCGCTCGTTGCTGGACTTAACTAGACACCTTGCGGCACGAGCT
GACGACGACCATGCACCACCTGTATAAAGTGTCCGATTGCTCGGACTAAGCCGCTTTCACGGCTGACACATTATATGTC
AAGTCTAGGTAAGATTCTTGGCTTGCTATCCAATTAATCTAGACAATCCACTGCTTGTGTGTTCCCCCGCCTATCCCTTT
GAGTTTCACTCTTGCGAGTATACTACCCAGGCGATCCGCTTAACGTGTTAACTTGCGGCACCCCACCCCGATAAATCGG
GGCAAGACACCTAGCGGACATAATTTACGGTTTGGACTACCAGGGTATCTAATCCTGTTTGCTCCCCAAACTTTCGTCTC
TCACCGTCAGGTTTATTGTAGTTAGCTGCTTTCGCTTTCGGCATTCTTTCCGATATCAACACATTTCACCGCTCCACCGG
AAATTCCACTAACCTCCAACAACCTCTAGTTAAATAGTATCTATTCCCGCTCCTCGATTAAGTCGAGGCTTTTAAAAACA
GACTTATTTATCAGGCTACAGACGCTTTACGCCCAGTAAATCCGGATAACGCTTGGGACCTTCGTATTACCGCGGCTGC
TGGCACGAAGTTAGCCGTCCCTTATTCTTGGTGTACCGTCATTCGTCTTCCACCAAAAAAGAAGTTTACACCCCTTAGGG
CTTAATCCTCCACGCGGAATTGCTGGATCAGGCTTTCGCCCATTGTCCAATATTCCCCACTGCTGCCTTCCGTAGAAGTA
TGGACCGTTTCTCAGTTCCATTGCGGCCGGACATCCTCTCAGACCGGCTACCCGTCATAGGCTTGGTAGGCCATTACCCT
TACCAACTACCTGATAGGCCGCACCCTCCTCTTTGAGTGATAAATCTTTCCTCATCATATTTTCATTTTGATGAGAATAT
TGAGAATTAGCCATATCTCTATGGTTGTTCTCAGCTCAAAGGCAGATTAAGTACGTGTTACTCGCCCGTTTGCCACTAAC
ATAATCTTGTATTGCTACAAAACAATGTTCGTTCGACTTGCATGTGTTAAGCATTCCGCTAGCGTTCATCCTGAGCCAGG
ATCAAACTC

>AR2A-1380R-046

TGAGTGCAAGGAACAGGGACATATTCACGGTGACATAGCTGATTCACCATTACTAGCAATTCCAACTTCATGTAGTCGA
GTTCCAGACTACAATCCGAACTAGGGGTAGGTTTGGAGATTTGCTCCACCTCGCGGTCTTGCTTCTCATTGTTACTACCC
ATTGTATTGTCTATGAAGCCCTAGGTGTAAGGGGCATGAGGATCTGACGTCATCCTCTCCTTCCTCCTGGTTACCCAGGC

AGTCTCCTTAGACATTCTAACTAAGGACGAGGGTTGCGCTCGTTACTAGACTTAACTAGACACCTTGCGGCACGAGCTG
ACGACGACCATGCACCACCTGTGCAAAGCTTCGTATTGCTACGAATAATATACTTTCATATATGACACATTGCATGTCA
ATCCTAGGTAAGATTCTTGGCTTTGTATCCAATTAATCTAGACAATCCACTGCTTGTGTGTTCTCCCGCCTATCCCTTTGA
GTTTCATTCTTGCGAATATACTACCCAGGCGGCTTGCTTAACTCGTTAGATTACGGCACTCAGATCCCGAAGGATCTAA
ACACCTAGCAAGCATAATTTAAGGTTTGGACTACCGGGGTATCTAATCCCGTTTGCTCCCCAAACTTTCGTCCCTCACCG
TCAGATGTATTGTAGTTAGCTGCTTTCGCTTTCGGCGTTCTTTCGGATATCAACGCATTTCACCGCTCCACCCGAAATTC
CACTAACCTCCAAATATCTCTAGATATGCAGTATCTATTCCCTTCCCTCGATTGAGTCCAGGTATTTGAAAACAGACTTA
CATATCAGGCTACGGACACTTTACGCCCAGTAAATCCGGATAACGCTTGGGACCTACGTATTACCGCGGCTGCTGGCAC
GTAGTTAGCCGTCCCTTATTCCTGGTGTACCGTCATTCGTCTTCCACCAGAAAAGGAGTTTACACACCGTAATGCGTCAT
CCTCCACGCGGAATCGCTGCATCAGGCTTTCGCCCATTGTGCAAGATTCCCCACTGTTGCCTCCCGTAGGAGTATGGGC
CGTATCTCAGTCCCATCGCGGCCGGACATCATCTCAGACCGGCTACCCGTCATAGGCTTGGTGAGCCATTACCTCACCA
ACTACCTGATAGGACGCACCCCCCTCCAAAAGCGATAAATCTTTAATCTAAGCTAATTTCATCACTTAGATACTATTGA
GGATTAGCTACATCTCTGTAGTTGGCCTCAACTTCTGGGCAGGTTAAGTACGTGTTACTCGCTCGTTTGCCACTAGTTGC
TATCTGTATTGCTACAAACAGAAACCCGTTCGACTTGCATGTGTTAGGCGTTCCGCTAGCGTTCATCCTGAGCCAGGAT
CAAACTC

>AR1A-1380R-34

TGAGTGCAAGGAACAGGGACATATTCACGGCGACATAGCTGATTCGCCATTACTAGCAATTCCAACTTCATGTAGTCGA
GTTCCAGACTACAATCCGAACTAAGGGTAGGTTTAAAGATTTGCTCCACCTCGCGGTATTGCTTCTCGCTGTTTCTACCC
ATTGTATTGTCTATGAAGCCCTAGGTGTAAGGGGCATGAGGATCTGACGTCATCCTCCCCTTCCTCCTGGTTACCCAGGC
AGTCTCTTTAGACATTCTAACTAAAGACGAGGGTTGCGCTCGTTACTAGACTTAACTAGACACCTTGCGGCACGAGCTG
ACGACGACCATGCACCACCTGTGCAAAGCTCCGTATTGCTACGGATAATATACTTTCATATATGACACATTGCATGTCA
ATCCTAGGTAAGATTCTTGGCTTTTTATCCAATTAATCTAGACAATCCACTGCTTGTGTGTTCCCCCGCCTATCCCTTTGA
GTTTCATTCTTGCGAATATACTACCCAGGCGGCTTGCTTAACTCGTTAGATTACGGCACTCAAATCCCGAAGGACTTAA
ACACCTAGCAAGCATAATTTAAGGCTTGGACTACCGGGGTATCTAATCCCGTTTGCTACCCAAGCTTTCGTCTCTCACCG
TCAGATGTATTGTAGTTAGCTGCTTTCGCTTTCGGCGTTCTTTCCGATATCAACGCATTTCACCGCTCCACCGGAAACTC
CACTAACCCCCAAATATCTCTAGATATGCAGTATCTATCCCCTTCCCTCGATTGAGTCCAGGTATTTGAAAACAGACTTA
CATATCAGGCTACAGACGCTTTACGCCCAGTAAATCCGGATAACGCTTGGGACCTACGTATTACCGCGGCTGCTGGCAC
GTAGTTAGCCGTCCCTTATTCCTGGTGTACCGTCATTCGTCTTCCACCAGAAAAGGAGTTTACACACCGTAATGCGTCAT
CCTCCACGCGGAATCGCTGCATCAGGCTTTCGCCCATTGTGCAAGATTCCCCACTGCTGCCTCCCGTAGGAGTATGGGC
CGTATCTCAGTCCCATCGCGGCCGGACATCATCTCAGACCGGCTACCCGTCATAGGCTTGGTGAGCCATTACCTCACCA
ACTACCTGATAGGACGCACCCCCCTCCAAAAGCGATAAATCTTTAATCTAAGCTAATTTCATCACTTAGATACTATTGA
GGATTAGCTACATCTCTGTAGTTGGCCTCAACTTCTGGGCAGGTTAAGTACGTGTTACTCGCTCGTTTGCCACTTTTATT
ACTCTGTATTGCTACAAAATAACTCTCGTTCGACTTGCATGTGTTAGGCGTTCCGCTAGCGTTCATCCTGAGCCAGGATC
AAACTC

>AR2A-1380R-141

GAGTTTGATCCTGGCTCAGGATGAACGCTAGCGGAACGCCTAACACATGCAAGTCGAACGAGATTTATTTTGTAGCAAT
ACAGAATAATGAAAGTGGCAAACGAGCGAGTAACGCGTACCTAATCTGCCCAGAAGTTGGGGATAACTATGGAGACAT
AGCTAATCCCCAATAGCATCTAGATGACGAAATTAGTCTAGATTAAAGATTTATCGCCTTTGGAGGAGGGTGCGTCCTA
TCAGGTAGTTGGTGAGGTAATGGCTCACCAAGCCTATGACGGGTAGCCGGTCTGAGATGATGTCCGGCCGCGATGGGA
CTGAGATACGGCCCATACTCCTACGGGAGGCAGCAGTGGGGAATCTTGCACAATGGGCGAAAGCCTGATGCAGCGATT
CCGCGTGGAGGATGACGCATTACGGTGTGTAAACTCCTTTTCTGGTGGAAGACGAATGACGGTACACCAGGAATAAGG
GACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTCCCAAGCGTTATCCGGATTTACTGGGCGTAAAGAGTT
CGTAGCCTGATAAATAAGTCTGTTTTCAAATACCTCGACTCAATCGAGGAAAGGGGATAGATACTGTTTATCTAGAGAT
ATTTGGGGGTTAGTGGAATTTCGGGTGGAGCGGTGAAATGCGTTGATATCCGAAAGAACGCCGAAAGCGAAAGCAGCT
AACTACAATACATCTGACGGTGAGGAACGAAAGCTTGGGTAGCAAACGGGATTAGATACCCCGGTAGTCCAAGCCTTA
AATTATGCTTGCTAGGTGTCTAGACCCTTCGGGGTTTGGGTGCCGTAATCTAACGAGTTAAGCAAGCCGCCTGAGTAGT
ATATTCGCAAGAATGAAACTCAAAGGGATAGGCGGGGGAACACACAAGCAGTGGATTGTCTAGATTAATTGGATACAA
AGCCAAGAATCTTACCTAGATTTGACATGCATTATGTCTACAGTGAAAGCTGTATATTCGTAGCAATACGAAGTTTTGC
ACAGGTGGTGCATGGTCGTCGTCAGCTCGTGCCGCAAGGTGTCTAGTTAAGTCTAGTAACGAGCGCAACCCTCGTCCTT
AGTTAGAATGTCTAAGGAGACTGCCCGGGTAACCGGGAGGAAGGAGAGGATGACGTCAGATCCTCATGCCCCTTACAT
CTAGGGCTTCATAGACAATACAATGGGTAGTAACAACAAGAAGCAATACCGCGAGGTGGAGCAAATCTCTAAACCTAC
CCTTAGTTCGGATTGTAGTCTGGAACTCGACTACGTGAAGTTGGAATTGCTAGTAATGGCGAATCAGCTATGTCGCCGT
GAATATGTCCCTGTTCCTTGCACTCA

>AR2A-1380R-119

TGAGTGCAAGGAACAGGGACATATTCACGGCGACATTGCTGATTCGCCATTACTAGCAATTCCAACTTCATGTAGTCGA
GTTCCAGACTACAATCCGAACTGAGGGTAGGTTTCAGGTTTGGCTCCATCTTACGATATTGCTTCCCATTGTTTCTACCC
ATTGTATGATCTATGAAGCCCTAGACATAAGGAGCATGAGGATCTGACGTCATCCTCTCCTTCCTCTCCGTTACCGAAG
CAGTCTCGTGTGACATTCTAACACACGATGAGGGTTGCGCCCGTTACTAGAATTAACTAGACACCTTGCGGCACGAGCT
GACGACGACCATGCACTACCTGTGCAGAGTTATGTATTGCTACATATAATCAGCTTTCACTGATGACACGCTGCATGTC
AAGCCTAGGTAAGATTTTTGGCGTGCTATCCAATTAATCGAGATCATCCACTGCTTGTGTGTTCCCCCCGCCTATTCCTTT
GAGTTTCATTCTTGCGAATATACTACCCAGGTGGTTCGCTTAACTCGTTAGATTACGGCACTCCCGCGAAAACACGGAA
ACACCTAGCGAACATAATTTACAGCTTGGACTACCGGGGTATCTAATCCCGTTTGCTACCCAAGCTTTCGTCTCTCACCG
TCAGTAATATTGTAGTTAACTGTCTTCACTTTTGGCATTCCTTCTGATATCAACGCATTTCACCGCTCCACCAGAAATTC
CATTAACCCCCAATACACTCCAGTAATGCAGTATCCATTCCCTCTCCCTAGTTGAGCTAAGGTATTTGAAAACAGACTT
ACAAAACCGGCTACAGACTCTTTACACCCAGTAAATCCGGATAATGCTCGGGACCTACGTATTACCGCGGCCGCTGGC
ACGTAGTTAGCCGTCCCTTATTCCTGAGGTACCGTCATTCTTCGTCCCCCAGAAAAGAAGTTTACACCCCAAAAGGCTT
CGTCCTTCACGCGAAGTCGCTGCATCAGGCTTTCGCCCATTGTGCAAGATTCCCCACTGCTGCCTCCCGTAGGAGTATG

GACCGTATCTCAGTTCCATTGTGGCCGGACATCCTCTAAGACCGGCTACCCGTCATAGGCTTGGTGGGCTATTACTCCG
CCAACTACCTGATAGGCCACATCCCCCTCCAAAAGCGATAAATCTTTCTCCAGCAATAATTTCGTCTTGCTGGACACAT
CCGGAATTAGCCATATCTCTATGGTTGTGCCGAACTTAAGGGCAGGTTAGACATGTGTTACGCACCCGTTTGCTACTAA
TTTAATCTAGTATTGCTACCAAACCAAATCCGTTCAACTTGCATGTGTTAAGCACTTCGCTAGCATTCATCCTGAGCCAG
GATCAAACTC

>AR2A-1380R-205

GAGTTTGATCCTGGCTCAGGATGAATGCTAGCGAAGTGCTTAACACATGCAAGTCGAACGAATTTAATTTAATAGCAAT
ATTGAGTTAAATTAGTGGCAAACGGGTGCGTAACACATGTCTAATCTACCCTTAGGTTCGGTATAACTATAGAGATATA
GCTAATCCCGGATATGATTCAAGAAATGAAAATACTTGAATGAAAGATTTATCGCCTTTGGAGGAGGATGTGGCCTATC
AGGTAGTTGGTGGAGTAATAGCCCACCAAGCCTATGACGGGTAGCCGGTCTTAGAGGATGTCCGGCCACAATGGAACT
GAGACACGGTCCATACTCCTACGGGAGGCAGCAGTGGGGAATCTTGCACAATGGGCGAAAGCCTGATGCAGCGACTTC
GCGTGAAGGACGAAGCCTTTTGGGGTGTAAACTTCTTTTCTGGGGGACGAAGAATGACGGTACCTCAGGAATAAGGGA
CGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTCCCGAGCATTATCCGGATTTACTGGGTGTAAAGCGTCTGT
AGCCGGTTTTATAAGTCTGTTTTCAAATACCTTGGCTCAACCAAGGAGAGGGAATGGATACTGTATTACTGGAGTATAT
TGGGGGTTAATGGAATTTCTGGTGGAGCGGTGAAATGCGTTGATATCAGAAGGAATGCCAAAAGTGAAGACAGTTAAC
TACAATATTACTGACGGTGAGAGACGAAAGCTTGGGTAGCAAACGGGATTAGATACCCCGGTAGTCCAAGCTGTAAAT
TATGTTCGCTAGGTGTTTCTATGTTTTCATAGGAGTGCCGTAATCTAACGAGTTAAGCGAACCACCTGGGTAGTATATTC
GCAAGAATGAAACTCAAAGGAATAGGCGGGGGAACACACAAGCAGTGGACGATCTCGATTAATTGGATAACACGCCA
AAAATCTTACCTAGGCTTGACATACAGTGTGTCATCAGTGAAAGCTGGTTATTCGTAGCAATACGAAGCGCTGAACAGG
TAGTGCATGGTCGTCGTCAGCTCGTGCCGCAAGGTGTCTAGTTAATTCTAGTAACGGGCGCAACCCTCATCGTATGTTA
GAATGTCATACGAGACTGCTTCGGTAACGGAGAGGAAGGAGAGGATGACGTCAGATCCTCATGCTCCTTATGTCTAGG
GCTTCATAGATCATACAATGGATAGTAACAATAGGAAGCAATATCGTAAGATGGAGCCAAACCTAAAACCTATCCTCA
GTTCGGATTGTAGTCTGGAACTCGACTACATGAAGTTGGAATTGCTAGTAATGGCGAATCAGCAATGTCGCCGTGAATA
TGTCCCTGTTCCTTGCACTCA

>AR1A-1380R-65

TGAGTGCAAGGAACAGGGACATATTCACGGCGGTATGGCTGACCCGCCATTACTAGCAATTCCAACTTCATGTAGTCGA
GTTCCAGACTACAATCCGAACTGAGGATAGGTTTAATGTTTTGCTCCATCTCGCGATATTGCTTCACGCTGTCCTACCCA
TTGTATTATCTATGACGCCCTAGACATAAGGGGCATGAGGATCTGACATCATCCTCTCCTTCCTCTTCGTTACCGAAGCA
GTCTCGTATGACATTCTAACATACGACGAGGGTTGCGCTCGTTACTAGACTTAACTAGACACCTTGCGGCACGAGCTGA
CGACGACCATGCACCACCTGTACATAGCTTCCTATTGCTAGGAATAATATAATTTCTTATATGACACAATGTATGTCAA
GCCTAGGTAAGATTCTTGGCTTATTATCCAATTAATCTAGATAATCCACTGCTTGTGTGTTCCCCCGCCTATCCCCTTGA
GTTTCACTCTTGCGAGTATACTTCCCAGGCGGCTTGCTTAACCCGTTAGGTACGGCACTCAAACCGACGAATCGATTCA
AACACCTAGCAAGCATCATTTACGGCTTGGACTACCGGGGTATCTAATCCCGTTCGCTCCCCAAGCTTTCGTCTCTGACC

GTCAGTTATTTTATAGTTAGCTGCTTTCGCTTTCGGCATTCCTTCCGATATCAACGCATTTCACCGCTCCACCGGAAATTC
CACTAACCTCTAAAATCCTCTAGATATACAGTATCTATCCCTTTTCTCCGATTGAATCGGAGTATTTAAAGACAGACTTA
TACATCCGGCTGCAGACGCTTTACGCCCAGTAATTCCGGATAACGCTCGAGACCTACGTATTACCGCGGCTGCTGGCAC
GTAGTTAGCCGTCTCTTATTCCTGAGATACCGTCATTCGTCTTCTCCCAGAAAAGAAGTTTACACCCCACGAGGCTTCAT
CCTTCACGCAGAATCGCTGGATCAGGCTTTCGCCCATTGTCCAATATTCCCCACTGCTGCCTCCCGTAGGAGTATGGAC
CGTTTCTCAGTTCCATTGCGGCCGGACATCCTCTCAGACCGGCTACCCGTCATAGGCTTGGTGAGCCGTTACCTCACCA
ACTACCTGATAGGACGCATCCCCCTCCTTGAGCGATAAATCTTTCTCCTACATATTTTCATTTTGTAGGACGTATGAGGA
ATTAGCTATGTCTCCATAGTTTTGCCTCACTCAAGGGCAGGTTAGATACGCGTTACTCGCCCGTTTGCCACTAACGCATT
ATCCGAAGACAATGAATTCGTTCGACTTGCATGTGTTAAGCGTTCTGCTAGCGTTCATCCTGAGCCAGGATCAAACTC

>AR2A-1380R-032

TGAGTGCAAGGAACAGGGACATATTCACGGCGGTATGCTGACCCGCCATTACTAGCAATTCCGACTTCATGCAGTCGA
GTTCCAGACTGCAATCCGAACTTGGGGTAGCTTTGTAAGATTTGCTGAGTCTCGCGGCCTCGCTTCTCGCTGTACTACCC
ATTGTATTGTCTATGATGCCCTAAACGTAAGGGGCATGAGGATCTGACGTCATCCTCTCCTTCCTCTCCGTTGCCGGAGC
AGCCTTGTTAGACATTCTAACTAGCAACGAGGGTTGCGCTCGTTCCCAGACTTAACTGGACACCTTGCGGCACGAGCTG
ACGACGACCATGCACCACCTGTATACCGTATCCAGATTGCTCTGGACTACTACATTTTCACGTAGGACACAGTATATGT
CAAGTCTAGGTAAGATTCTTGGCTTCGTATCCAATTAATCGAGACAATCCACTGCTTGTGTGTTCCCCCGCCTATCCCTT
TGAGTTTCATTCTTGCGAATATACTACCCAGGCGACATGCTTAACTCGTTAGATTACGGCACCCCACCCACATAATGTG
GATAAGACACCTAGCATGCATAATTTACAGCTTGGACTACCGGGGTATCTAATCCCGTTTGCTCCCCAAGCTTTCGTCTC
TGACTGTCAGATTTACTATAGTTGGCTGCTTTCGCTTTTGGCGTTCCTTCCGATATCAACGCATTTCACCGCTCCACCGG
AAATTCCACCAACCCCTAGTAATCTCTAGATTTGCAGTATCCATTCCCTTTCTATGGTTAAGCCACAGGATTTGAAAATA
GACTTATAAATCCAGCTACAGACTCTTTACGCCCAGTGATTCCGGATAACACTTGGGACCTACGTATTACCGCGGCTGC
TGGCACGTAGTTAGCCGTCCCTTATTCCTGGTGTACCGTCATTCGTCTTCCACCAGAAAAGAAGTTTACACACCATAATG
CTTCATCCTTCACGCGGGATTGCTGCGTCAGGCTTTCGCCCATTGCGCAAGATTCCCCACTGCTGCCTCCCGTAGGAGTA
TGGACCGTGTCTCAGTTCCATTGCGGCCGGACATCCTCTCAGACCGGCTACCCGTCATAGGCTTGGTAGGCCATTACCC
CACCAACTACCTGATAGGACGCACCCCCCTCCAAAAGCGATAAATCTTTGATCTAAAAGCATTTCAGCTTCAAGATAAT
ATCGAGGATTAGCTACGTCTCCGTAGTTGTCCTCAACTTCTGGGTAGGTTAAGTACGCGTTACTCGCCCGTTTGCCACTC
TCCGATTGCTCGGAGCGTTCGACTTGCATGTGTTAAGCATCCCGCTAGCGTTCATCCTGAGCCAGGATCAAACTC

>AR2A-1380R-022

TGAGTGCAAGGAACAGGGACATATTCACGGCGGTGTGGCTGACCCGCCATTACTAGCAATTCCAACTTCATGTAGTCGA
GTTCCAGACTACAATCCGAACTGAGGATAGGTTTCGCGTTTTGCTCCATCTCACGATATTGCCTCGCATTGTCCTACCCA
TTGTATTATCTATGACGCCCTAGACATAAGGGGCATGAGGATCTGACATCATCCTCTCCTTCCTCTTTGTTACCAAAGCA
GTCTCCTGTGACATAATAACACAGGACGAGGGTTGCGCTCGTTGCTAGACTTAACTAGACACCTTGCGGCACGAGCTGA
CGACGACCATGCACCACCTGTACAACGCTTCGTATTGCTACGAATAGTAATGTTTCCACTACGACACAATGTATGTCAA

GCCTAGGTAAGATTTTTGGCTTATTATCCAATTAATCTAGATAATCCACTGCTTGTGTGTTCCCCCGCCTATCCCTTTGA
GTTTCACTCTTGCGAGTATACTTCCCAGGCGACTTGCTTAATCCGTTAGGTTCGGCACTCAATCCGCAGAAGCGAATCA
AACACCTAGCAAGCATCATTTACGGCTTGGACTACCGGGGTATCTAATCCCGTTTGCTACCCAAGCTTTCGTCTCTGACC
GTCAGATATTTTATAGTTAGCTGCTTTCGCTTTCGGCATTCCTTCCGATATCAACACATTTCACCGCTCCACCGGAAATT
CCACTAACCTCTAAAATCCTCTAGGTATACAGTATCTATCCCTTTTCTCAGATTGAATCAGAGTATTTAAAGACAGACTT
ATACACCCGGCTGCAGACGCTTTACGCCCAGTAATTCCGGATAACGCTCGAGACCTACGTATTACCGCGGCTGCTGGCA
CGTAGTTAGCCGTCTCTTATTCCTGAGGTACCGTCATTCGTCTTCCCCCAGAAAAGAAGTTTACACCCCACGAGGCTTCA
TCCTTCACGCAGAATCGCTGGATCAGGCTTTCGCCCATTGTCCAATATTCCCCACTGCTGCCTCCCGTAGGAGTATGGAC
CGTTTCTCAGTTCCATTGCGGCCGGACATCCTCTCAGACCGGCTACCCGTCATAGACTTGGTGAGCTATTACCTCACCAA
CTATCTGATAGGCCGCATCCCCTTCCTTGAGCGATAAATCTTTCCTCCACATATTTTCATTTTGTGGAGCGTATGAGGAA
TTAGCTACATCTCTGTAGTTGTGCCTCGCTCAAGGGCAGGTTAGATACGTGTTACGCGCCCGTTTGCCACTAACTTGTAT
TGCTACAAATTCGTTCGACTTGCATGTGTTAAGCGTTCTGCTAGCGTTCATCCTGAGCCAGGATCAAACTC

>AR2A-1380R-015

TGAGTGCAAGGAACAGGGACATATTCACGGCGACATTGCTGATTCGCCATTACTAGCAATTCCAACTTCATGTAGTCGA
GTTCCAGACTACAATCCGAACTGGGGGACGGTTTCGCGTTTTGCTCCACCTCGCGGCATTGCTTCGCATTGTCCGTCCCA
TTGTATGACCTATGAAGCCCTAGACATAAGGGGCATGAGGATCTCACGTCGTCCTCCCCTTCCTCCCTGTTGCCAGGGC
AGTCTCGTGTGACATACTAACACACGACGAGGGTTGCGTCCGTTTCTAGAATTAACTAGACACCTTGCGGCACGAACTG
ACGAGGACCATGCACCACCTGCACAACGCTTTGTATTGCTACAAATAACCGGCTTTCACCGATGACACATTGCATGTCA
AGCCTAGGTAAGATTGTTGGCTTGCTATCCAATTAATCTAGATCATCCACTGCTTGTGTGTTCCCCCGCCTATCCCTTTG
AGTTTCACTCTTGCGAGTATACTACCCAGGCGGTTCGCTTAACCCGTTAGGTTCCAGCACTCCCCAGGTATAACCTGAG
AAACACTTAGCGAACATAGTTTACGGCGTGGACTACCGGGGTATCTAATCCCGTTTGCTCCCCACGCTGTCGTCTCTCA
CCGTCAGTAAGATATGGTTAGCTGCTTTCGCTTTTGGCATTCCTTCCTATATCAACGCATTTCACCGCTCCACAGGAAAT
TCTACTAACCCCTATCATACTCTAGGTATGCAGTATCTATTCCCTTTTCCCAATTGAGTTGAGATATTTGAAAATAGACT
TACAAACCAGGCTACAGGCGCTTTACGCCCAGTAACTCCGGATAACGCTTGGGACCTTCGTATTACCGCGGCTGCTGGC
ACGAAGTTAGCCGTCCCTTATTCCTGAGGTACCGTCATTCGTCTTCCCCCAGAAAAGAAGTTTACACCCCACGAGGCTT
CATCCTTCACGCAGAATCGCTGCATCAGGCTTTCGCCCATTGTGCAATATTCCCCACTGCTGCCTCCCGTAGGAGTATGG
ACCGTGTCTCAGTTCCATTGCGGCCGGACATCCTCTCAGACCGGCTACCCGTCATAGGCTTGGTAGGCCATTACCCCAC
CAACTACCTGATAGGCCGCATCCCCGTCCCAAAGCGGTAAACCTTTGCTTCACATATTTTCATTTTGTGAGGCGCATCCT
GGATTAGCCATATCTCTATGGTTGGCCAGAACTTCGGGGAAGGTTAGATACGTGTTACGCGCCCGTTTGCCACTATCAT
GTATTGCTACATGACCGTTCGACTTGCATGTGTTAGGCGTTCTGCTAGCGTTCATCCTGAGCCAGGATCAAACTC

>AR2A-1380R-165

GAGTTTGATCCTGGCTCAGGATGAACGCTAGCGGGATGCTTAACACATGCAAGTCAAACGGCCAGCAATGGCAGTGGC
AAACGGGCGAGTAACGCGTACCTAACCTACCCAGAAGTTGGGGACAACTACGGAGACGTAGCTAATCCCCAATACTCC

CTACACAATGAAAAAATGTAGGGCAAAGATTTATCGCTTTTGGAGGGGGGTGCGTCCTATCAGGTTGTTGGTGGGGTA
ATGGCCTACCAAGCCTATGACGGGTAGCCGGTCTGAGAGGATGTCCGGCCGCAATGGAACTGAGACACGGTCCATACT
CCTACGGGAGGCAGCAGTGGGGAATTTTGCGCAATGGGCGGAAGCCTGACGCAGCAATCCCGCGTGAAGGATGAAGC
ATTATGGTGTGTAAACTTCTTTTCTGGTGGAAGACGAATGACGGTACACCAAGAATAAGGGACGGCTAACTACGTGCC
AGCAGCCGCGGTAATACGTAGGTCCCAAGTGTTATCCGGAATTACTGGGCGTAAAGAGTCTGTAGCCTGACGCATAAG
TCTGTTTTCAAATCTTGTGGCTTAACCACAAAAAGGGAATGGATACTGTGTGTCTGGAGATTACTAGGGGTTGGTGGAA
TTTCCGGTGGAGCGGTGAAATGCGTTGATATCGGAAGGAACGCCAAAAGCGAAAGCAGCCAACTATAGTAAATCTGAC
GGTCAGAGACGAAAGCTTGGGGAGCAAACGGGATTAGATACCCCGGTAGTCCAAGCTGTAAATTATGCATGCTAGGTG
TCTTATCCACTTTTGTGGGTGGGGTGCCGTAATCTAACGAGTTAAGCATGTCGCCTGGGTAGTATATTCGCAAGAATGA
AACTCAAAGGGATAGGCGGGGGAACACACAAGCAGTGGATTGTCTCGATTAATTGGATACGAAGCCAAGAATCTTACC
TAGACTTGACATATACTGTGTCCTACGCGAAAGTGTAGTAGTCCATAGCAATATGGACACGGTATACAGGTGGTGCATG
GTCGTCGTCAGCTCGTGCCGCAAGGTGTCCAGTTAAGTCTGGGAACGAGCGCAACCCTCGTTGTTAGTTAGAATGTCTA
ACAAGACTGCTCCGGCAACGGAGAGGAAGGAGAGGATGACGTCAGATCCTCATGCCCCTTACGTCTAGGGCATCATAG
ACAATACAATGGGTAGTACAGCGAGAAGCGAAACCGCGAGGTCGAGCAAATCTTACAAAGCTACCCCTAGTTCGGATT
GCAGTCTGGAACTCGACTGCATGAAGTCGGAATTGCTAGTAATGGCGGGTCAGCATACCGCCGTGAATATGTCCCTGTT
CCTTGCACTCA

>AR1A-1380R-33

TGCAGTCGAACGAGTTTTATTTTGTAGCAATACAGAATAAAATTAGTGGCAAACGAGCGAGTAACGCGTACTTAATCTG
CCCAAAAGTTGAGAACAACTATAGAGATGTAGCTAATTCTCAATATGCTCACGATGGTGAAAATCCGATGTGAGGAAA
GATTTATCGCTTTTGGAGGAGGGTGCGTCCTGTCAGGTTGTTGGTGAGGTAATAGCTCACCAAGCCTATGACGGGTAGC
CGGTCTGAGATGATGTCCGGCCGCGATGGAACTGAGATACGGCCCATACTCCTACGGGAGGCAGCAGGGGGGAATCTT
GCACAGGGGGAAAAAGCCTGATGCAGCTTTTCCGCGGGGGATGTGACATTTTAGGGTGTAAACTCCTTTTTTGGGGGAA
GAAGAATGACGGTACTCCAGGAATAAGGGACGGCTAATTACTTGCCAGCAGCCGCGGTAATACGTAGGTCCCAAGCGT
TATCCGGATTTACTGGGCGTAAAGTGTCCGTAGCCTGGTATATAAGTCTGTTTTCAAATACCTCGACTCAATCGTGGAA
AGGGAATAGATACTGTATACCTAGAGAATATTGGAGGTTAGTGGAATTTCCGGTGGAGCGGTGAAATGCGTTGATATC
GGAAAGAACGCCGAAAGCGAAAGCAGCTAACTATGATATTTCTGACGGTGATGGACGACAGCTTGAGTAGCAAACGG
ATTAGATACCCCGGTAGTTCAAGCCGTAAATTATCCTTGCTAGGTGTCTCCATAACTATTGGTATAGCATTTGTTATATC
AAAAGTTATGGAGGTGCCGTAAGTTAACACGTTAAGCAAGGCGCCTGTGTAGTATATTCGCAAGAATGAAACACAAAG
GGATAGGCGGGGGAACACACAAGCAGTGGATGATCTAGATTAATTGGATACTAAGCCAAGAATCTTACCCAGACTTGA
CATGGATTGTGTCTACGATGAAAGTCGTATAGCTATAGCAATATAGCGCTTTTCACAGATGGTGCATGGTCGTCGTCAG
CTCGTGCCGCAAGGTGTCTAGTTAAGTCTAGTAACGAGCGCAACCCTCATCGTTAGTTAGTTATGTCTAACGAGACTGC
CCGCGTAAGCGGGAGGAAGGAGAGGATGACGTCAGATCCTCATGCCTCTTACGTCTGGGGCTTCATAGATCATACAAT
GGGTAGTTACAATGAGAAGCAATACCGCAAGGTGGAGCAAATCTTCAAAGCTACCCTCAGTTCGGATTGTAGTCTGGA

ATTCGACTACATGAAGTTGGAATTGCTAGTAATGGTGAATCAGCCATGTCACCGTGAATATGTCCCTGTTCCTTGCACTC
A

>AR1A-1380R-72

TGAGTGCAAGGAACAGGGACCTATTCACGGCGGCTTGGCTGATCCGCCATTACTAGCGATTCCAACTTCATGTGGTCGA
GTTCCAGACCACAATCCGCACTTGGGATGGCTTTCGCGTTTCGCTCCTCCTCTCGGTATCGCTTCGTTCTGTACCACCCA
TTGTATTGCCTCTGACGCCCTGATCATAAGGGCCATAATGACCTGACATCATCCCCTCCTTCCTCCTCGTTGCCGAGGCA
GTCCACCTAGACATCCTAACTAAGTGTAAGGGTTGCGCTCGTTCCCGGACTTAACCGGACACCTTGCGGCACGAGCTGA
CGACGGCCATGCAGCACCTGCACAGCGCGATATATTGCTATATCTAGTATGACCTATCATACGACACACTGCATGTCAA
GATCAGGTAAGATTCTTGGCTTATTATCCAATTAATCGAGACAATCCACCGCTTGTGTGTTCCCCCGCCTATCCCTTTGA
GTTTCGGTCTTGCGACTATACTACTCAGGCGGCTCGCTTAACCCGTTTGGTTTCAGCACCCCCAATTACTCAGAGACACC
TAGCGAGCATAGTTTACAGTTTGGACTACCAGGGTATCTAATCCTGTTTGCTACCCAAACTTTCGTCCTTTACCGTCAGA
TAAGGCATAGCCAGCTGCTTTCGCTTTCGGCATTCCTTCCGATATCAACAGATTCCACCCCTCCACCGGAAATTCTACTA
ACCTCCACCTATCTCTAGCTTAACAGTATCTGCCCCCTCTCCGAAGTTGAGCCCCGGTTTTTAAAAGCAGACTTATTATG
CAGGCTACAGACTCTTTACGCCCAGTGATTCCGGATAATGCTCGAGACCTACGTATTACCGCGGCTGCTGGCACGTAGT
TAGCCGTCTCTTATTCCTGGTGTACCGTCATTCGTCTTTCACCAGAAAAGAAGTTTACACCCCTTAAGGCTTCGTCCTTC
ACGCGGAGTTGCTGGATCAGGGTTCCCCCCATTGTCCAATATTCCCCACTGCTGCCCCCCGTAGGGGTATGGGCCGTGT
CTCAGTCCCATTGCGGCCGGTCACCCTCTCAGGCCGGCTATCCGTCTTAGTCTTGGTAAGCTCTTACCTCACCAACTAAC
TGGTGGTCCGCACCCCCCTCCCAGAGCGATAAATCTTTCTCCAGATCTTTCTCATTCCTCCAGACGTATACAGAATTAGC
TGCGTTTCCGCAGTTGTCCTCTACTCCAGGGCAGGTTAGGCACGTGTTACTCGCCCGTTCGCCACTGCCATTACTGGCCG
TTCGACTTGCATGTGTTAAGCACTCCGCTAGCATTCATCCTGAGCCAGGATCAAACTC

>AR2A-1380R-222

GGCAAACGAGCAAGTATCGCGTACTTAGTCTGCCCAGAAGTTGAAACTCTCTATGCAGATGTACCTAATTCTCAATATG
CTCCCAATGGTGAAAATCCNATGTGGATAAATATTTATCCCTTTTGGAGGAGGGTGCGTCCTATCAGGTTGTTGGTGAG
GTAATANCTCACCAATGATATGACGGGTGGCCGGTCTGATATGATGTCCGGCCGCAATGGAACTGAAATACGGACCAT
ACTCCTACGGGGGGCAGCGGAGGGNTGTCTTGCGCATGGGGCGAAGCCCTGATGCTTCGATTCCCCGTGTAGGATGAC
GCGTTACGGTGTGTAAACTCCTTTTTTGAGGAAAACGAATGATGGTGCACCATAGGATAGGGACGGTTACCTTCCTGCC
ACCACCCGCGGAATTACGTAGGTCCCGCTCTTTATCCGGATTTACTGGCCGTAAAGTGCCTTTGCCTGGTTTATTGCTTC
TGTTTTCAAATGCCGCGACTCCATCGTGGAAAGGGAATAGATACTGTATACCTAGAGAATATTGGAGGTTAGTGGAATT
TCCGGTGGAGCGGTGAAATGCGTTGATATCGGAAAGAACGCCGAAAGCGAAAGCAGCTAACTATGATATTTCTGACGG
TGATGGACGACAGCTTGAGTAGCAAACGGGATTAGATACCCCGGTAGTTCAAGCCGTAAATTATCCTTGCTAGGTGTCT
CCATAACTATTGGTATAGCATTTGTTATATCAAAAGTTATGGAGGTGCCGTAAGTTAACACGTTAAGCAAGGCGCCTGT
GTAGTATATTCGCAAGAATGAAACACAAAGGGATAGGCGGGGGAACACACAAGCAGTGGATGATCTAGATTAATTGG
ATACTAAGCCAAGAATCTTACCCAGACTTGACATGGATTGTGTCTACGATGAAAGTCGTATAGCTATAGCAATATAGCG

CTTTTCACAGATGGTGCATGGTCGTCGTCAGCTCGTGCCGCAAGGTGTCTAGTTAAGTCTAGTAACGAGCGCAACCCTC
ATCGTTAGTTAGTTATGTCTAACGAGACTGCCCGCGTAAGCGGGAGGAAGGAGAGGATGACGTCAGATCCTCATGCCT
CTTACGTCTGGGGCTTCATAGATCATACAATGGGTAGTTACAATGAGAAGCAATACCGCAAGGTGGAGCAAATCTTCA
AAGCTACCCTCAGTTCGGATTGTAGTCTGGAATTCGACTACATGAAGTTGGAATTGCTAGTAATGGTGAATCAGCCATG
TCACCGTGAATATGTCCCTGTTCCTTGCACTCA

>AR2A-1380R-112

NNNNNNNNGGGCGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGGCTT
TGAGTGCAAGGAACAGGGACATATTCACGGTGACATGGCTGATTCACCATTACTAGCAATTCCAACTTCATGTAGTCGA
ATTCCAGACTACAATCCGAACTGAGGGTAGCTTTGAAGATTTGCTCCACCTTGCGGTATTGCTTCTCATTGTAACTACCC
ATTGTATGATCTATGAAGCCCCAGACGTAAGAGGCATGAGGATCTGACGTCATCCTCCCCTTCCTCCCGCTTACGCGGG
CAGTCTCGTTAGACATAACTAACTAACGATGAGGGTTGCGCTCGTTACTAGACTTAACTAGACACCTTGCGGCACGAGC
TGACGACGACCATGCACCATCTGTGAAAAGCGCTATATTGCTATAGCTATACGACTTTCATCGTAGACACAATCCATGT
CAAGTCTGGGTAAGATTCTTGGCTTAGTATCCAATTAATCTAGATCATCCACTGCTTGTGTGTTCCCCCGCCTATCCCTT
TGTGTTTCATTCTTGCGAATATACTACACAGGCGCCTTGCTTAACGTGTTAACTTACGGCACCTCCATAACTTTTGATAT
AACAAATGCTATACCAATAGTTATGGAGACACCTAGCAAGGATAATTTACGGCTTGAACTACCGGGGTATCTAATCCCG
TTTGCTACTCAAGCTGTCGTCCATCACCGTCAGAAATATCATAGTTAGCTGCTTTCGCTTTCGGCGTTCTTTCCAATATC
AACGCATTTCACCGCTCCACCGGAAATTCCACTAACCTCCAATATTCTCTAGGTATACAGTATCTATTCCCTTTCCACGA
TTGAGTCGTGGCATTTGAAAACAGACTTATATACCAGGCTACGGACACTTTACGCCCAGTAAATCCGGATAACGCTTGG
NACCTACGTATTNCCGNGGCTGCTANNNACGNNNNNNNNCGTNNNN

>AR2A-1380R-64

GNNNTTNNGTGACCTATAGAATACTCAAGCTATGCATCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCC
AGTGTGCTGGAATTCGGCTTTGAGTGCAAGGAACAGGGACATATTCACGGYGGCATGCTGAYYCGCCATTACTAGCAA
TTCCRACTTCATGYAGTCGARTTCCAGACTRCAATCCGAACTGAGGGTAGCTTTGAAGATTTGCTCGACCTYGCGGTWT
CGCTTCTCRYTGTACTACCCATTGTATKRTCTATGAWGCCCYAGACGTAAGRGGCATGAGGATCTGACGTCATCCTCTC
CTTCCTCYCSSTTRCSSGRGCAGTCTYGTTAGACATWACTAACTAACGAYGAGGGTTGCGCTCGTTMCYAGACTTAACT
AGACACCTTGCGGCACGAGCTGACGACGACCATGCACCAYCTGTRWAMMGYGTCYATATTGCTATRGACTACTACACT
TTCRTCGTAGGACACARTMTATGTCAAGTCTRGGTAAGATTCTTGGCTTMGTATCCAATTAATCTAGAYCATCCACTGC
TTGTGTGTTCCCCCGCCTATCCCTTTGAGTTTCATTCTTGCGAATATACTACCCAGGCGCCTTGCTTAACKYGTTAAMTT
ACGGCACCCCATACCCTTGAATACAAAAGCATAATGGTTAGGAGACACCTAGCATGSATAATTTACGGCTTGGACTACC
GGGGTATCTAATCCCGTTTGCTCCCCAAGCTTTCGTCTCTGACCGTCAGAATTACCATAGTTGGCTGCTTTCGCTTTCGG
CGTTCCTTCCGATATCAACGCATTTCACCGCTCCACCGGAAATTCCACCAACCCCTAATATTCTCCAGATACACAGTATC
CATTCCCTTTTTGTGGTTGAGCCACGAGATTTGAAAACAGACTTATGCGTCAGGCTACGGACTCTTTACGCCCAGTAATT
TCNTAACGCTTNNNCNNNNTTATTTACGCGGNCNNNN